

Time for a belated rethink of fusion's goals

The ITER partners must acknowledge that the world has changed since the project's inception, and reconsider what they want from collaboration on magnetic confinement fusion.

At its inception in 1985, the International Thermonuclear Experimental Reactor (ITER) was a geopolitical construct, an agreement made by the presidents of the United States and the Soviet Union to nurture mutual trust. The project soon became a four-way partnership, involving Europe and Japan. But when the Soviet Union ceased to be, it soon became apparent that Russia would not contribute equally to ITER's costs. Subsequently, the United States also decided not to contribute its full share.

The two partners left with money on the table ignored America's 1995 suggestion that ITER be reconfigured, claiming that they could carry the bulk of the estimated \$10 billion costs of ITER construction. Over the past year, France, Germany and finally Japan have looked the costs of hosting ITER in the eye, and blinked (see pages 115–119).

The context of the ITER project has changed beyond all recognition in the twelve years since its inception. With the passing of the Cold War, the need to engage the United States and Russia in scientific and industrial collaboration has regrettably slipped off the agenda of presidential summits. The willingness of governments to invest in energy research has also faded. Work on renewable energy sources and on the efficient use of fossil fuels has been scaled back, and research into better fission reactors, including fast reactors, has been mothballed.

And rather than achieving a consensus on the worth of the tokamak concept encapsulated in ITER, fusion researchers have become more divided on the proposed design and its ability to meet its various objectives. They have also been unable to dispel the chronic scepticism of some nuclear engineers and materials scientists about the feasibility of the tokamak as the basis for any future power reactor.

Neither ITER nor its member governments have proved responsive to these changes. ITER is a diffuse beast, with no major commercial contractor in charge to fight its corner in the corridors of power. The governments have neither signalled clear support for ITER construction nor prepared alternatives in the event that construction

cannot go ahead next year, as it was supposed to do.

Those governments have three choices now. They can agree to extend the Engineering Design Assessment for a few years, in the (probably forlorn) hope that money to build ITER will become available later on. This seems to be the favoured option at the moment.

They can withdraw from ITER, and concentrate whatever money they have left on theoretical work and on alternatives to the tokamak. But this could consign magnetic fusion to a kind of paralysis. Plasma will never be successfully confined by a computer.

Or they can reconfigure the collaboration with a view to building a more affordable experiment, perhaps one that can sustain the burning of plasma for a set period of time. But there is no certainty that the smaller amount of money needed to build such an experiment would be available once its design was completed.

The harnessing of fusion energy is a worthy goal, and fusion science has a fine and honourable record of international collaboration. But the partners now appear to be working together for the benefit of the collaboration itself without sufficient consideration of changing circumstances. It is imperative that ministers of the governments involved get together for a frank exchange of views on what they wish to achieve through their collaboration on magnetic fusion. Do they want a new fusion science experiment, for example, or a test-bed for fusion reactor technology? ITER wants to be both, but it is not affordable. Or would they be better off pursuing fusion science at a sustainable level, while addressing some of the engineering and materials science issues that would have to be overcome before any real technology demonstrator of magnetic confinement fusion could be built?

Confidence in such a reassessment will require technically informed advice from outside the fusion community about which avenues are worth pursuing, and strong will from the partner governments to bring new impetus to the collaboration. Regrettably, neither commodity is readily available as the 1998 construction decision looms. □

Tremors that should be public

The verification of the CTBT will provide data whose open availability would serve scientists and defence interests.

Few partnerships can have been as successful as the one that has existed, for the past forty years, between seismology and monitoring of underground nuclear tests. That the world now has a signed Comprehensive Test Ban Treaty (CTBT) is due in no small part to the past successes of seismological monitoring. And in turn, a generation of US seismologists has received generous funding from the Air Force for fundamental research.

With the signing of the CTBT, the science of test-ban monitoring is set to enter a new era. There should be a wealth of new data — from seismometers, but also from new networks of hydrophones, atmospheric pressure gauges and radionuclide detectors. Alas, there is no commitment by the US defence department to make these data avail-

able to the international scientific community (see page 107).

This state of affairs is worrying not just for research scientists but for the test-ban treaty itself. The problem of effective test-ban verification is far from solved, and the best solution to a difficult problem is to let as many bright people attack it as possible. Moreover, while openness may sometimes be inconvenient, a treaty such as the CTBT depends on mutual confidence, which can only be enhanced by free access to the data on which verification depends.

Thanks in part to the efforts of Vice President Al Gore, recent years have seen some welcome releases of previously classified datasets to the research community (see, for example, *Nature* 379, 300–301; 1996). The signatories to the CTBT should not buck the trend. □



Loss of Japanese satellite deals blow to remote sensing efforts

[TOKYO] Officials at Japan's National Space Development Agency (NASDA) were both surprised and disappointed last week when the agency's US\$1.2-billion environmental monitoring satellite ADEOS — the Advanced Earth Observation Satellite — ran out of power and stopped working.

The satellite, launched less than a year ago, carried sensors not only from Japan but also from the United States and Europe for monitoring changes in the global environment. It is the first of a series of such large satellites planned by the agency.

ADEOS stopped sending signals during the morning of Monday, 30 June. It was subsequently discovered that the satellite's solar panel had ceased to function, and that ADEOS had switched to a low-energy-consumption mode. No immediate cause for the breakdown could be identified, and after several failed attempts to reactivate the satellite, NASDA made the breakdown public.

At a meeting of the Space Advisory Committee the following day, several explanations for the failure of the satellite's power system were suggested, including collision with meteorites or 'space junk'. But there is increasing evidence that the failure was due to a design flaw in the tension-adjustment mechanism of the satellite's solar panel.

According to Tasuku Tanaka, director of NASDA's Earth Observation Research Center, the value of the data gathered during ADEOS's short mission is not yet clear. But

he admits that the seven and a half months of actual observation time — less than a quarter of the scheduled 33 months — may be insufficient to answer the global issues ADEOS was supposed to address.

As one success of ADEOS, Tanaka cites results on Arctic ozone depletion, gathered by a total ozone mapping spectrometer developed by the US National Aeronautics and Space Administration (NASA), which point to a phenomenon similar to the Antarctic 'ozone hole'.

These results appear to have been confirmed by data from the Improved Limb Atmospheric Spectrometer (ILAS) on ADEOS, developed by the Japanese Environmental Agency, which measured the vertical distribution of ozone and various greenhouse gases in the atmosphere. But researchers at the National Institute for Environmental Studies say that the data obtained will now have to be supplemented by ground-based measurements.

Other equipment on board ADEOS included a sensor for measuring the Earth's radiation budget, developed by the French space agency CNES. NASDA officials say they are "surprised" by the accident. But they are confident the loss will not affect the overall development of remote sensing and environmental monitoring capabilities at the agency.

As the first in a series of international environmental monitoring satellites sched-

uled for launch over the next few years, ADEOS is the cornerstone of an ambitious strategy at NASDA to develop indigenous capabilities in remote sensing. Two more environmental monitoring satellites — the Tropical Rainfall Measuring Mission and a successor to ADEOS — are scheduled for launch by 2000.

NASDA's environmental monitoring programme, which consumes almost 25 per cent of Japan's total space budget, is the cornerstone of research on global change in Japan. While universities spend only about ¥1.3 billion (US\$11 million) annually on global environmental research, NASDA's remote-sensing programme consumes almost ¥40 billion. As a consequence, any changes in NASDA's remote-sensing programme would badly affect the state of global-change research in the country, says Akimasa Sumi, a professor of climatology at the University of Tokyo.

Mike Mann, a deputy associate administrator of NASA, described the loss of ADEOS as "a real blow to NASA's science programme". Mann said in a statement that, fortunately, much of the data from ozone monitoring instruments aboard ADEOS could be replaced by those from instruments on other spacecraft. "But the sea-surface wind data provided by the NASA Scatterometer will be harder to replace, and were opening essentially new opportunities for research and operational users worldwide." **Robert Triendl**

After 322 years, royal observatory loses out to Scottish rival



Northern lights: The Royal Observatory in Edinburgh, which is to become the seat of a new National Astronomy Technology Centre.

[LONDON] The Royal Greenwich Observatory (RGO), currently based in Cambridge, is to close and merge with the Royal Observatory in Edinburgh, with the loss of up to a hundred jobs from both sites. But British astronomy will gain from savings of at least £2.4 million (US\$4 million) a year for the next four years, and £4 million a year thereafter. The merged site will be renamed the Astronomy Technology Centre.

The decision ends 15 years of uncertainty over the future of the two royal observatories. It was greeted with relief by the Particle Physics and Astronomy Research Council (PPARC), which runs the two sites. The move is expected to take place in July 1998.

But staff at the RGO, as well as the Astronomer Royal, Sir Martin Rees, say they are "devastated" and will "fight the decision all the way". They believe the decision is a huge loss to British science and a blow to Britain's international standing in astrono-

my. Rees, who campaigned for both sites to remain open, says the savings will be heavily outweighed by the loss of both an historic observatory and an "excellent" modern scientific institution.

A survey of Britain's astronomers conducted by the Royal Astronomical Society earlier this year, however, showed that most believed one of the two sites had to go. This was also the view of the directors of both sites, and the society's president, Malcolm Longair, believes the decision was the right one.

Ken Pounds, chief executive of PPARC, says the savings come at a time when British astronomy is going through one of the most cash-starved periods it has known. If the merger had not happened, he says, the research council would have been compelled to make savage cuts to grants to university astronomy research programmes.

Meanwhile the name of the Royal Greenwich Observatory, which was founded in

1675, will almost certainly remain, and could return to its original site next to the Thames in south London, where its former building is now a museum.

The decision to close the Cambridge site was announced last week by the science minister, John Battle, and was taken on the advice of PPARC (see *Nature* 387, 646; 1997). Pounds says the council could not afford to maintain two institutions essentially performing the same function — providing technical support to Britain's telescopes in the Canary Islands and Hawaii.

PPARC's predecessor, the Science and Engineering Research Council, had wanted to merge the observatories at Edinburgh 15 years ago. But this was considered too politically sensitive at the time, says Pounds. A later panel voted to move the RGO to Cambridge in 1990 on the strength of possible collaboration between the RGO and the University of Cambridge's highly regarded astronomy facilities.

In 1995 a review panel on UK millimetre, optical and infrared astronomy again decided to merge the centres. That review was chaired by James Hough, head of the department of physical sciences at the University of Hertfordshire. But its plans were put on hold when the previous Conservative government launched its 'prior options' initiative inviting competitive bids from the private sector to manage UK research facilities.

A new panel was convened after the May general election, this time chaired by Brian Eyre, deputy chairman of AEA Technology plc. This panel unanimously came to essentially the same conclusion. Eyre says there was little to choose between the two sites. Edinburgh was chosen because it offered the right mix of skills for PPARC programmes.

But RGO staff strongly disagree. David Carter, who runs the RGO's telescope design consultancy jointly with Liverpool John Moores University, says Edinburgh does not have the same expertise in telescope design as

Cambridge. He fears these skills will be lost overseas. He doubts whether staff will want to relocate to Edinburgh given the recent move to Cambridge from the RGO's former home in East Sussex.

Andy Lawrence, professor of astronomy at the University of Edinburgh and a member of the management board at the Royal Observatory in Edinburgh, agrees that Cambridge possesses superior design skills. But he says that PPARC chose Edinburgh because its superior instrumentation skills will be more useful now that the United Kingdom's last big telescope project that needed design input, the twin 8-metre Gemini, is nearing completion.

The RGO left its Greenwich site 50 years ago, moving to East Sussex after the Second World War to escape the streetlights and smog of London, then to Cambridge. It won the race to fix longitude at sea, established the meridian, and set Greenwich Mean Time as the international standard. **Ehsan Masood**

'Political interference skewed scientific advice on fish stocks'

[MONTREAL] Three Canadian scientists claim that political and bureaucratic interference in fisheries science has compromised the government's efforts to sustain stocks of Atlantic cod and Pacific salmon.

Jeffrey A. Hutchings of Dalhousie University's biology department, Carl Walters of the University of British Columbia's fisheries centre, and Richard L. Haedrich of the biology department at Memorial University of Newfoundland claim that the administrative framework linking science with management suppresses scientific uncertainty and obscures scientists' differences of opinion.

They propose replacing it with a politically independent organization of fisheries scientists. They also suggest that all scientific information about fish stocks should be released to the public at the same time as it is presented to the fisheries department, so that the public can evaluate management decisions based on that information.

But the scientists' ideas have been dismissed by officials from the fisheries department. The department's deputy director, William Rowat, claims that the comments are based on innuendo and misrepresentation, and are part of a vendetta against the department, its scientists and its managers.

The scientists' arguments appeared last month in the *Canadian Journal of Fisheries and Aquatic Sciences*, published by the National Research Council of Canada, under the title "Is scientific inquiry incompatible with government information control?"

To back up their claims that a political 'spin' is being placed on scientific results, the authors refer to several incidents in which they allege that government fisheries reports excluded scientific information contrary to



the official line. They claim that the government, which denies that overfishing is the primary cause of present stock collapses, omitted references to conclusions that overfishing had caused stock decline in a 1995 report for Newfoundland groundfish.

Scientific information was also selectively excluded in the 1995 Stock Status Report on Gulf of St Lawrence groundfish, say the scientists. The original draft of the document said that seal predation or environmental conditions were unlikely to be responsible for cod mortality trends from 1985–87. But this statement was removed from the published version, contrary to scientific advice, the authors claim.

They also allege that scientists have been ordered not to discuss politically sensitive matters — such as overfishing — in public, "irrespective of the scientific basis, and publication status, of the scientist's concerns".

One scientist who admitted in an interview with a journalist in 1995 that east coast fish stocks had collapsed from overfishing

and "had nothing to do with the environment, nothing to do with seals" — as some fishermen had claimed — was officially reprimanded for not giving a balanced perspective and for disagreeing with the Newfoundland Stock Status Report. Yet "these comments were consistent with much of the research that had been ... published in peer-reviewed journals", the authors say.

The authors claim that inappropriate government influence on fisheries science also extended to testimony given by scientists in the courts. They quote one scientist who described his confusion when told how to behave as an expert witness in a case involving salmon affected by a dam built by the aluminium smelting company Alcan.

The scientist wrote in 1986 that at the meeting the director-general in the fisheries department had instructed staff to support the minister's position, while adhering to the scientific advice. "I find it impossible to do both," he wrote.

William Doubleday, director-general, science, in the fisheries department, has criticized the comments as "not the usual scientific debate" but "an attack on an organization and the people that were working in it".

Doubleday says the article contains "factual errors, misrepresentations, and very selective quotes". He says the department is preparing a rebuttal, but that the authors have also been invited to participate in an open forum to debate the management of science later this summer.

But the former editor of the journal, David Cook, who gave up the post last month, defended publication of the article by suggesting it would lead to "broad exposure" and "candid debate". **David Spurgeon**

NRC backs open access to test-ban data

[WASHINGTON] Scientific researchers should have open access to data collected for monitoring the Comprehensive Test Ban Treaty, says a panel of the US National Research Council (NRC), the operating arm of the National Academy of Sciences. The panel also calls for a substantial increase in funding for research to improve test-ban monitoring.

Although the treaty states that its provisions “shall not be interpreted as restricting the international exchange of data for scientific purposes”, concerns were raised about open access to the monitoring data during the negotiation of the treaty in Geneva. According to members of the NRC panel, sensitivity is likely to centre on radionuclide data, which might be used by environmental groups to identify leaks from nuclear power plants.

Some countries are also thought to have expressed a more general concern that making the monitoring data widely available will allow outside bodies — even the news media — to draw premature conclusions about potential treaty violations, before the parties to the treaty have had a chance to analyse the data fully.

But Thorne Lay, of the Institute of Tectonics at the University of California, Santa Cruz, and chair of the NRC panel, dismisses the “irrational concern that CNN might scoop the national data center”. Lay says that media organizations lack the necessary seismic data-processing capability but that, if they had it, “they could do this right now” using publicly available seismograms.

The test-ban treaty has been signed by 142 countries including the five nuclear weapons states (the United States, Russia, China, France and the United Kingdom). It calls for the setting-up of an International Monitoring System, consisting of seismic, hydroacoustic, infrasound and radionuclide sensors around the world, and reporting back to an International Data Centre in Vienna.

About 10 gigabytes of data per day will reach the centre, which will then distribute information to data centres in each of the countries participating in the treaty. It has been left to the treaty’s preparatory commission, currently meeting in Vienna, to draft a policy on the distribution of data beyond the national data centres.

The US government has yet to formulate a policy on this question. Steven Bratt, principal programme director for nuclear treaty programmes in the office of the Secretary of Defense, says that the United States plans “to have a position on this by the time it comes up in Vienna”, but the issue is “not at the top of the queue”.

In the meantime, however, the Air Force Technical Applications Center at Patrick Air Force Base in Florida, which has been providing a prototype US national data centre for an international test of the seismic component of the treaty monitoring system, has told US seismologists that, for the time being, it will not provide open access to data from stations outside the United States. “We can’t make a decision about other countries’ data,” says Bratt.

The NRC panel’s report, which was published last week, argues that open, near-real-time access to the data is essential to support the United States’ ability to monitor the treaty and to allow for use of the data in other fields of basic and applied research.

For example, the 60-station infrasound network should provide important information on volcanic eruptions in remote areas, as well as helping to detect meteors entering the Earth’s atmosphere. And the seismic data, if available in near real-time, will help in the rapid location and size estimation of earthquakes — information that could help to accelerate emergency response efforts.

Despite statements from defence department officials supporting open data access, seismologists in the academic community have expressed scepticism about the intention of the US national data centre to distribute data freely. “This was a classified arena,” says one NRC panel member. “They aren’t used to people looking [over their shoulders].”

The potential for difficult relations between the defence and academic communities has also arisen in the context of the funding of basic research to support the test-ban treaty. The NRC report outlines the research necessary to support the United States’ test-ban monitoring goals, and concludes that substantially increased funding will be needed. Its rationale is that “there are major unsolved problems in seismology, and that there will soon be a substantial flow of data from [the other monitoring] systems for which there is far less operational experience”.

But funding has decreased, from about \$12 million per year in fiscal years 1995 and 1996 to \$9.2 million in the present fiscal year. At the same time, the defence department’s share of the funding (\$8.8 million), previously split into separate lines for basic, exploratory-development and advanced-development research, has been consolidated into a single line controlled by the Nuclear Treaty Program Office (NTPO). This has prompted worries in the academic community that, as one panel member put it, “calibrating seismometers will suck up basic research money”.

A more immediate concern is that the NTPO, through its administrative agent the Defense Special Weapons Agency, has apparently managed to spend only \$3.8 million of the \$8.8 million allotted to it. One seismologist suggests that the agency “was not used to dealing with basic research and the academic community. They’re used to dealing with transfers to national labs, not small (\$50,000–200,000) grants.”

Steven Bratt of the NTPO admits: “We didn’t do as well as we should have this year. Next year we’re going to try to do better.”

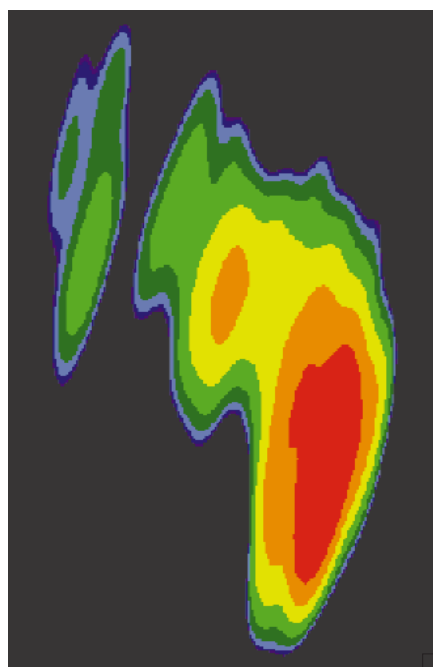
Laura Garwin

First radiotelescope images from space

[LONDON] The first images obtained using Very Long Baseline Interferometry (VLBI) with a Japanese radioastronomy satellite were released last week. The image of the quasar 1156+295 (right) is therefore one of the first images to have been produced using a radiotelescope in space.

Researchers at the National Radio Astronomy Observatory in Socorro, New Mexico, combined signals from the Japanese HALCA satellite and from ground-based stations in the National Science Foundation’s Very Long Baseline Array (VLBA) and Very Large Array (VLA). The VLBA has 10 antennas and a maximum separation of 20 miles, and the VLA has 27 antennas and a maximum separation of 5,000 miles.

HALCA, launched in February by Japan’s Institute of Space and Astronautical Science, is the first satellite designed for radio-astronomy imaging. This image shows the quasar’s core, bottom right, and a jet of subatomic particles emerging from the core towards the top left.



El Niño forecast fails to convince sceptics

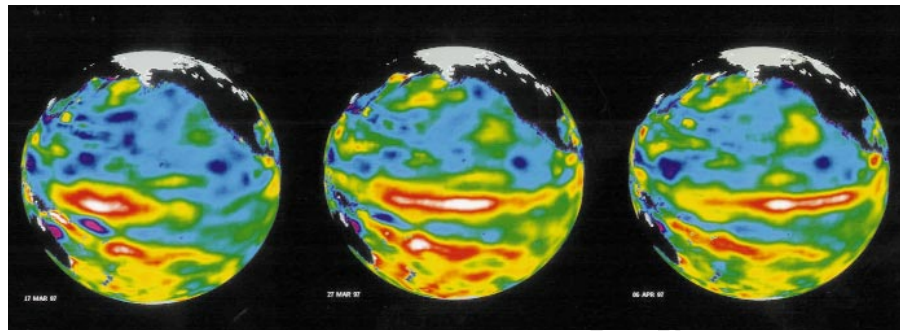
[LONDON] Predictions that the expected El Niño — the appearance of warm waters off the coast of South America, bringing droughts, storms and floods in its wake — will be stronger than usual are facing scepticism from the countries that could be most strongly affected.

In particular, scientists from Argentina and Peru say they are not yet convinced by reports from the US National Oceanic and Atmospheric Administration (NOAA) that El Niño, which usually occurs in winter every two to seven years, could arrive early, and be the strongest for 25 years.

The forecast, issued on 26 June, says that Indonesia and India are expected to be drier than normal (see panel), while central Chile, Uruguay, southern Brazil, Peru and central Argentina are likely to experience a wetter than normal winter.

But Pablo Lagos, scientific director at the Geophysical Institute of Peru in Lima, is unconvinced. He says it is far too early to take drastic steps, such as changing agriculture patterns, in response to the NOAA prediction. "What is the point of spending so much money if the prediction turns out to be wrong?" he says. "It is better to wait and act only when we are sure."

Hugo Hordij, director of the Climatic Analysis Centre of Argentina's National Meteorological Service, agrees, adding that



These images of sea-surface height relative to normal ocean conditions are evidence of an imminent 'bad' El Niño. A warm water oscillation (white/red) is shown in the equatorial Pacific during March and April this year. The data were collected by the joint US/French TOPEX/Poseidon satellite.

he doubts Argentina will be affected by El Niño to the extent predicted by NOAA.

The NOAA prediction was based partly on the abnormally high sea-surface temperatures that have been recorded in parts of the East Pacific, and along and near the west coast of South America. These temperatures continued to increase throughout June, and are now around 2–3°C higher than average.

Such conditions are similar to those observed in 1957 and 1972, both years in which an El Niño event led to droughts near the eastern end of the Pacific Ocean and caused increased rain leading to floods at the western end of the Pacific.

Michael Glantz, a senior scientist at the

National Center for Atmospheric Research at Boulder, Colorado, and author of *Currents of Change: El Niño's Impact on Climate and Society*, believes that the advice given to the governments of Peru and Argentina by their scientists is prudent. "You can't take these predictions at face value," he says.

"There is no one-to-one correlation between high sea-surface temperatures and an El Niño," says Glantz. For example, he says, the 1991–92 El Niño was preceded by variable temperatures off the coast of Peru which fluctuated considerably above and below their average values.

Lagos, meanwhile, says that Peru is not wholly disregarding the latest forecast, merely waiting for more precise data. Peru, he says, has a comprehensive two-stage action plan in place, should an El Niño strike. The plan is based in part on the lessons that the country learned in 1982–83, when a lack of preparation for what turned out to be the strongest El Niño since 1950 led to 600 deaths as a result of severe flooding, and US\$900 million worth of damage.

The first phase, which Lagos says has already been set in motion, consists of campaigns to make the public aware of the possibility and the possible impact of an El Niño. The second phase starts "once we are more sure about an El Niño", Lagos adds.

This phase includes drilling into underground water aquifers to ensure clean water supplies, and encouraging the textile industry to ensure that stocks of warm clothing are protected from floods. A law is being considered banning anchovy fishing to conserve depleted fish stocks, which migrate during an El Niño. Farmers, meanwhile, may be encouraged to grow crops that would benefit from a wetter climate.

An El Niño, however, is not all bad news for Peruvian farmers. Some take advantage of the prospect of torrential rainfall by planting copious quantities of seed in desert areas, in the hope of cultivating pastures for their cattle. "Many farmers have already started doing this," says Lagos.

Ehsan Masood

Indian monsoon 'progressing normally'

[NEW DELHI] The Indian Meteorological Department has expressed doubts about whether the predicted return of a strong El Niño this year will cause drought in some parts of India and affect the seasonal monsoons.

"The southwest Indian monsoon arrived a week late but, contrary to fears, it has so far been progressing normally and has already covered most of the country," says R. R. Kelkar, a senior official of the department.

Kelkar was reacting to statements by Ants Leetmaa, director of the Climate Prediction Center of the US National Oceanic and Atmospheric Administration, that the 'big' El Niño predicted this year for the equatorial Pacific is likely to bring drought to India, South Africa, Australia and parts of Brazil (see above).

According to Indian scientists, India's monsoon is "a gigantic complex phenomenon" in which El Niño is just one of several contributory factors. In the monsoon season, which runs from June to September, about 80 cm of rain falls throughout India. The department believes that India will get around 92 per cent of that amount this year, which is considered satisfactory.

This prediction is based on an empirical 16-parameter model that the department has been using successfully since 1988 (see *Nature* 342, 4; 1989). Kelkar says that the El Niño phenomenon has been incorporated into calculations for this year's forecast.

After analysing monsoon rainfall data between 1951 and 1990, Indian

meteorologists say that there is no clear correlation between El Niño and the Indian monsoon. "There have been many years in the past when monsoon rainfall was in excess or normal in El Niño years, and there were years when rainfall was deficient when there was no El Niño," says Kelkar. Indeed, India had a normal monsoon in 1983 when Peru witnessed its worst El Niño since 1950.

India's monsoons over the past nine years have proved unexceptional, and the US warning has come as something of a surprise to the government. Although the meteorological department claims that it has allayed fears, the agriculture ministry is preparing to install pumps and borewells for irrigating crops, in case the rains fail.

K. S. Jayaraman

EU helps to save science jobs in Berlin

[MUNICH] Sixty east Berlin scientists who had kept their jobs thanks to now-defunct 'rescue' funds following the reunification of Germany have been saved from unemployment once again by a new scheme supported by the European Union (EU).

The DM18.6-million (US\$10.6-million) scheme, which began last week, will retrain the scientists, most of them at the Max Delbrück Centre for Molecular Medicine (MDC). The money is from the European Social Fund, part of the EU's 'structural funds' which subsidize its poorer regions.

If successful, the scheme could become a model for unemployed researchers in other poor regions in east Germany and elsewhere in Europe. Brandenburg, for example, is already preparing a similar proposal.

After reunification in 1990, many of the estimated 24,000 scientists from the former East Germany had difficulty finding jobs in the new research organizations that were introduced, modelled on those of West Germany. Almost 75 per cent lost their jobs.

For a few years, several thousand of the most highly qualified scientists stayed in work thanks to special support programmes, such as the Wissenschaftler-Integrations-Programm (WIP) and the Verstärkungsfonds. But these ran out at the end of last year, and the German government has refused to make further money available (see *Nature* 378, 656; 1995). In Berlin alone, over 2,000 scientists joined the dole queues in January.

One of the problems for east German scientists has been that their lack of experience in modern scientific techniques has made it difficult for them to compete for grant money with those from the west, according to Marion Bimmler of the MDC's personnel department.

"Neither German programmes, nor the EU's Framework research programmes, offer adequate retraining opportunities for researchers with a slightly outdated scientific background," she says. Two years ago, she decided to try to secure money from the European Social Fund — which, among other objectives, supports retraining of the unemployed — for the east German scientists in her institute.

But these funds are normally intended for retraining those with few, if any, qualifications. To make qualified researchers eligible, Berlin had to adapt its three-year plan for spending its current allocation of money from the social funds, and then gain the approval of Germany's 15 other states.

Final approval now depends on the European Commission being prepared to modify its normal rules, although it is widely expected that it will do so. A formal decision, which the MDC is confident will be in its favour, is due in the next few weeks.

"We are very enthusiastic about the idea of using social funds in this way," says a commission official, "particularly because the part of the social funds which we have earmarked for spending on research activities is consistently being underspent, even though we consider research to be a central point in regional development."

The commission is anxious that social funds used to retrain researchers should have a long-term impact for the region, rather than being just a short-term welfare act. So it will be closely watching the experiment in Berlin to ensure that scientists find appropriate jobs after retraining. "In the long term this would help to establish industrial R&D in east Germany," says the official. But he warns that the social funds must not simply become a covert means of paying researchers' salaries.

The 60 Berlin researchers will work on reduced salaries during the three years. Unlike normal social funds training programmes, which are full-time, the training of the east German scientists in the theory and practice of molecular biology will be limited to eight hours a week. During the rest of the

time they will be active members of MDC research groups. "Being given the chance to improve their scientific competence will make it easier for the scientists to be more successful in applying for grant money afterwards," says Bimmler.

Bimmler's hopes of placing all 160 newly out-of-work MDC scientists on this scheme could not be realized, but those who will benefit are relieved and optimistic. For example, Karla Köpke, a 46-year-old biologist who was out of work for a year after reunification, and who has since survived on two WIP grants at the Humboldt University, will be trained in human genetics at the MDC, joining the Human Genome Project research group.

Having worked as a pharmacologist for 20 years in the former East Germany, then in protein design and biometrics at the Humboldt, Köpke has no experience in human genetics or modern molecular biological techniques. "Given my age and background," she says, "I have no chance of finding a job on the regular market." After retraining she hopes to win grants to allow her to continue working at the MDC.

Quirin Schiermeier

Tariff waiver would aid instrument imports

[WASHINGTON] Draft legislation has been introduced into the US Congress to remove a longstanding impediment to international collaboration by lifting tariff restrictions on pieces of large scientific instruments brought into the United States from other countries.

The bill was introduced last month by James Sensenbrenner (Republican, Wisconsin), chairman of the US House of Representatives Science Committee.

Projects affected include one of two Gemini telescopes now under construction in Hawaii. The US Customs Service originally ruled that a mirror, manufactured in the United States but sent to France for polishing, could not re-enter the country without duty being paid. But Congress passed legislation last year to waive the tariff, saving Gemini \$1.8 million, according to Richard Malow of the Association of Universities for Research in Astronomy, the consortium managing the project.

Similar problems have faced the Large Binocular Telescope project in Arizona and the Department of Energy's Continuous Electron Beam Accelerator Facility in Newport News, Virginia.

Sensenbrenner's bill is designed to amend the language of the Florence Agreement governing the import of items of educational, scientific or cultural benefit. Duty-free status would extend not just to whole instruments but also to components, when "the instrument or apparatus, due to



Sensenbrenner: lowering costs.

its size and complexity, cannot be imported in its assembled state". The Florence Agreement dates back to the 1960s. According to one Customs Service official, "They didn't foresee these huge instruments you couldn't possibly bring in whole."

The current regulation clearly exempts components from duty-free status, so the only way to solve the problem is with new legislation. But the Commerce Department would still have to rule that there is no US equivalent for the part being imported.

Within days of introducing the bill, Sensenbrenner received a congratulatory letter from Burton Richter, director of the Stanford Linear Accelerator Center in California, saying the legislation would help smooth collaboration on the centre's B-factory particle collider. International partners are picking up some 40 per cent of the cost of the instrument. "They, and we, have been very worried about the possibility that tariffs will be assessed on their contribution, which would raise the cost of the detector significantly," Richter wrote.

The bill has been referred to the House Ways and Means Committee for consideration, and Malow expects it eventually to pass.

Tony Reichhardt

CNRS returns fire on ethics review panel

[PARIS] Institutional arrangements within France's public research agencies often encourage the politicization of decision-making, as well as nepotism and cronyism, according to an ethics committee set up by the Centre National de la Recherche Scientifique (CNRS), Europe's largest basic-research agency.

These conclusions are contained in a critical report, *Ethics and Scientific Institutions*, released in Paris last week. But the report has already sparked controversy over the role of the committee itself, with some senior science administrators arguing that questions about the organization of research are beyond its remit and competence.

The committee, known as COMETS, is chaired by Hélène Ahrweiler, a historian and president of the Paris-based University of Europe. It was set up to identify ethical problems in areas other than biology and medicine (see *Nature* 370, 88; 1994). Its 14 members include Hubert Curien, a former research minister, and Jean Dausset, 1980 Nobel prizewinner in physiology or medicine.

Its report emphasizes that, overall, the French research system works well. But,

acknowledging that it is focusing on 'black spots', the report is particularly critical of the research organizations' executive boards. These bodies are made up of public personalities, leaders of industry and representatives of the government and trade unions, and in principle wield considerable power, as they approve all budgetary decisions, and define the agencies' research strategy.

The report says the boards of the research agencies are failing to exercise their powers of oversight, in particular in the choice of research strategy. Real power tends to be concentrated in the hands of the offices of the director-generals of the research agencies and the research ministry, according to the report, which claims that the boards often do little more than rubber-stamp decisions.

The boards of agencies such as CNRS should speak out on strategic issues, says the report. It claims that many major decisions are taken behind closed doors, allowing lobby groups to flourish, in particular in defining 'big science' projects. Edouard Brezin, president of the CNRS board, concedes that it is "a pity we are not consulted on strategy before decisions are taken; we are often asked to vote

afterwards simply as a legal requirement".

To remedy the situation, the report suggests that boards be given greater independence, pointing out that the representatives of the ministry on boards currently have a right of veto, which sharply curtails the boards' autonomy.

The report also criticizes the procedure used to appoint senior staff. The director-generals of research agencies are nominated directly by government. They then nominate other senior staff directly in consultation with the ministry. The report recommends that the executive boards of the agencies should set up search committees to propose candidates for appointments.

The lack of such committees is the "French tragedy *par excellence*", asserts one member of the ethics committee, arguing that this results in excessively politicization of nominations for senior posts. "The power to nominate is decisive in France; by nominating people we make them our debtors," he adds.

Meanwhile, the release of the report has sparked a row over the role of the ethics committee itself. The three-year mandate of current members of COMETS expired last month, and CNRS has now decided to review the committee's functions, one idea being to replace it with a national committee. CNRS is holding discussions with the research ministry, and a decision on the committee's future is expected at the autumn meeting of the CNRS board.

Brezin argues that the committee's report overstepped its remit, and that its contents concern organizational matters that have "nothing to do" with ethics. While agreeing that there are "elements of truth" in the report, he claims that its conclusions are too simplistic, and that the committee's members lack sufficient "competence" to address such issues. Brezin says he would be open to an audit of "legitimate questions" concerning the running of French research, on condition that this "be done seriously by competent people".

One member of COMETS claims that the decision to restructure the committee stems in part from the fact that the report's criticisms have not been appreciated by senior CNRS management. CNRS's action, he claims, is a thinly veiled warning to their successors on whatever form of committee emerges that "they don't have the right to make too severe criticisms".

Brezin dismisses these claims, arguing that the decision to review the status of the committee at the expiry of the current members' mandates was logical, particularly given the arrival of the new government. "There is no intention to abolish or suspend the committee," he says.

Declan Butler

US budget plans jeopardize role in ITER

[WASHINGTON] A key congressional budget committee will this week consider fresh and drastic cuts in fusion research which could lead to the abrupt withdrawal of the United States from the International Thermonuclear Experimental Reactor (ITER) project.

Early drafts of the appropriations bills for the 1998 financial year (starting in October), to be finalized on 11 July by the energy and water appropriations subcommittee in the House of Representatives, would cut the Department of Energy's fusion research programme from \$225 million to \$175 million, according to administration officials.

The proposal would take most of the \$50 million cut from the US contribution of \$55 million to ITER. That would effectively end US participation in the engineering design assessment phase of the project nine months before its completion next July (see Briefing, on page 115).

One administration official says that the corresponding Senate subcommittee is considering a similar cut, but many fusion supporters expect the Senate to be more supportive than the House.

Rumours of the proposed cut stunned the fusion community, which experienced deep budget cuts in 1995. "This is still a rumour, but we are taking it very seriously," says Anne Davies, head of the fusion programme at the energy department. "I'm shocked that

this could even be considered."

Jack Gibbons, science adviser to President Bill Clinton, was this week making contact with Joe McDade (Republican, Pennsylvania), chair of the House subcommittee, and Pete Domenici (Republican, New Mexico), chair of the Senate panel, to appeal for the administration's \$225 million request for fusion research to be met.

Staff on both subcommittees declined to comment in detail on their plans, but confirmed that they may pay particular attention to ITER because its construction has been delayed by the international partners (see *Nature* 387, 746; 1997).

In theory, the numbers drawn up by the staff have little significance until they are considered by the subcommittees. But in practice it can be difficult for members to find money to restore cuts proposed by staff.

Administration officials will argue that the United States has a commitment to its partners — Europe, Japan and Russia — to complete its share of ITER research. But this did not count for much in 1995, when the Congress cut the US contribution from \$80 million to \$55 million. Fusion lobbyists worry that few members of the appropriations panels are sufficiently committed to fusion, or to ITER, to find \$50 million from other programmes to support it.

Colin Macilwain

Mentors seek to prove value for money

[WASHINGTON & NEW YORK] Three years ago, the US Congress ended a \$15-million-a-year scheme by the Department of Energy to forge links between schoolteachers and scientists in its laboratories, after a report from the General Accounting Office (GAO) concluded that teaching science teachers about science was a waste of time and money.

This summer, hundreds of scientists will serve as mentors for schoolteachers who will spend their long vacation rediscovering science. Some are also seeking elusive data to prove the GAO wrong, and that teaching teachers about real science is an efficient way of raising school standards in science and mathematics.

In its report, GAO, the investigatory arm of Congress, attacked the energy department for concentrating resources on 'teacher enhancement' when, it said, "research suggests that these projects may be ineffective at increasing student achievement". But advocates of summer programmes for teachers — including some of the United States' most prominent scientists — believe fervently that they do just that. The challenge is to produce the data to prove it.

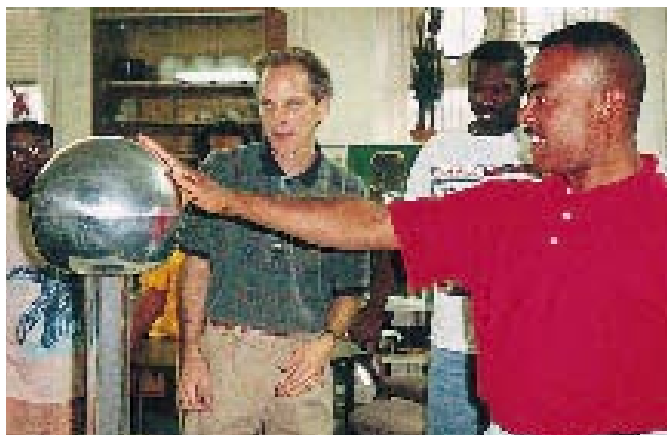
Washington's Carnegie Institution has pioneered efforts to support schoolteachers. When Maxine Singer, then head of the biochemistry laboratory at the National Cancer Institute, became president of the institution in 1988, she quickly sought to open up its building to children from local schools. It was parents at the nearby Ross Elementary school who suggested that she should train the teachers as well.

This year, 100 elementary schoolteachers will spend six weeks at the institution during the summer under what is now a major programme funded by the National Science Foundation (NSF), called the Carnegie Academy for Science Education (CASE).

Project-based experiments

Singer says the institution focused on elementary schools, which teach 5–11-year-olds, because "by the time you get to middle school, most of the inner-city kids are already convinced that science and maths are not for them". The non-specialist teachers go to class with three science educators who have been hand-picked by Singer to run the programme, as well as a larger number of teachers who excelled at CASE in previous years and who are returning as mentors.

They learn project-based approaches to simple experimental problems—finding out what determines the period of a pendulum, for example. And they are taught to link their teaching approach to the National Science Education Standards produced last year by the National Research Council — favouring discovery over the repetition of facts.



Hands-on experience: Charles James (left) is helping the Carnegie Institution in Washington to work with local schoolteachers under the sponsorship of the National Science Foundation during the long vacation to restimulate an active interest in and knowledge of science.

Alida James Fenner, one of the mentors and a teacher at Bunker Hill Elementary School in northeast Washington, says the course has entirely changed her approach to teaching science, and has led to its integration with other subjects, such as geography.

By the time it ends next year, 450 teachers will have attended CASE — a quarter of the elementary-school staff in the District of Columbia (DC) school system, one of the most troubled in the United States.

Rediscovering real science

Until recently, Carnegie had made little progress towards formal cooperation or support from the DC school board. But last year, Congress asked General Julius Beckton, a former Army commander, to take over responsibility for the system from the widely criticized board. Beckton has visited Carnegie several times, honoured the programme with a special award, and is now talking with Singer about a deal under which Carnegie will take responsibility for science and maths teacher training in the city's elementary schools.

About 90 smaller but more intensive programmes across the country are retraining high-school science teachers by giving them laboratory work experience. At Columbia University in New York, for example, ten teachers from local high schools are selected each year for two successive summers of hands-on experience in a research laboratory.

"This has been a very, very uplifting experience," says Edwin Klibaner, an experienced biology teacher at the John Dewey High School on Coney Island, South Brooklyn, who went through the scheme three years ago. "It reinforced the notion that I had the ability to do science — real science."

Klibaner now runs a biotechnology programme at John Dewey High. The Nobel laureate Joshua Lederberg of Rockefeller University was among those who helped Klibaner to design the programme. "I could not have had that kind of support and entry into the world of science without the

programme," says Klibaner. "And if I didn't do this, no one else at the school would."

New York's school system is ten times the size of Washington's, and the city school board has shown little interest in supporting this kind of intensive professional development for its teachers. But every student undergoes annual standardized tests, called Regent's tests, providing one of the largest and most complete databases of student performance in the United States.

For Sam Silverstein, chair of physiology at Columbia's medical school, these data represent a unique opportunity for formally evaluating the scheme against the only yardstick that matters to politicians — student test performance. Early assessments have shown positive trends in both attendance and Regent's test scores; but statisticians at Columbia point out that the number of teachers in the initial sample is too small to produce statistically significant results.

But does it improve teaching?

Silverstein, in collaboration with the organizers of five similar schemes across the United States, has applied to the NSF for a \$1.3-million grant to assess the schemes' effectiveness. He has already obtained some NSF money to develop the proposal.

GAO's 1994 report criticized the Department of Energy for the lack of formal studies on its programme's impact. Congress then shut down the programme, withdrawing \$15 million of annual support — about the same as is now spent on such schemes by all other sources combined. These programmes are not cheap to run. Columbia's costs \$25,000 per teacher over two years — half of it in teachers' stipends. But each teacher has hundreds of students, and an annual investment of \$75 million would allow every science teacher in the country to reconnect with science every thirteen years. Teachers love to learn science, but it will take hard data on student results to open the way for investment on that scale.

Colin Macilwain

Marine researchers in Europe 'must pool resources'

[MUNICH] Europe's marine scientists are calling for greater cooperation between national marine research organizations and industry to address important issues that individual organizations cannot afford to tackle on their own.

The scientists are proposing a scientific plan ranging from study of the deep sub-seafloor biosphere to marine drilling technologies. They are urging the European Commission to include at least part of the plan in its fifth Framework research programme, the detailed content of which is currently being discussed.

The proposals come in a position paper published last week by the marine and polar sciences board of the European Science Foundation. The board represents most marine science organizations in Europe, and the paper says that Europe has a world lead in some areas of exploration sciences, such as the study of continental margins and high-resolution studies of climate and sea-level changes.

But the report says that these strengths can be maintained and expanded only by pooling resources. It stresses that European scientists need more access to deep drilling

facilities, and says that Europe could effectively participate in the development of deep penetration exploration tools, such as new logging sensors, fluid samplers and in-situ laboratories.

AIDS 'cure' researchers guilty of misconduct

[CAPE TOWN] Two researchers at the University of Pretoria who claimed to have discovered a cure for AIDS in the form of the drug Virodene were last week found guilty of misconduct by a university disciplinary committee (see *Nature* 385, 474; 1997). The researchers, Dirk du Plessis, head of the university's department of cardiothoracic surgery, and Kallie Landauer, a consultant, were "sternly reprimanded" for proceeding with clinical trials of the drug after having been refused permission by the university's ethics committee. But both will retain their posts, according to Mike Smuts, a university spokesman.

Infrastructure shortfall hits UK universities

[LONDON] British universities faced a funding shortfall of £440 million (US\$700 million) for research between 1994–95 and 1995–96, according to a report commissioned by the Committee of Vice-Chancellors and

Principals, and the Higher Education Funding Council for England.

The study, by the management consultancy Segal Quince Wicksteed, is the latest document to conclude that income for research infrastructure has failed to keep up with income for research projects. Diana Warwick, the committee's chief executive, warned that the research reputation of British universities could be threatened unless a solution is found.

EU expected to appeal over beef hormones

[PARIS] The European Commission is expected to appeal against a ruling last week by the dispute panel of the World Trade Organization (WTO) that its ban on imports of beef from cattle treated with hormones was contrary to fair trade. The commission argues that the panel failed properly to consider the scientific arguments of its case, and that the ruling was contrary to the principle that countries have the right to fix their own levels of consumer protection.

The commission must win the approval of its member states before a formal appeal can be submitted. If the European Union loses the appeal, which is likely to be held later this year, it would be faced with the choice of either lifting the ban, or paying compensation to the United States for lost

trade. If it refused to do either, the WTO could authorize the United States to take retaliatory trade measures.

Russian pickets call for funding increase

[MOSCOW] More than 200 scientists picketed government buildings in Moscow last week demanding improved funding for Russian science and expressing concern that many of their jobs may be lost in forthcoming reforms. The scientists urged the government to provide at least 1,600 billion roubles (US\$300 million) a month for scientific research — twice the current level.

The scientists demanded a doubling of the salaries of all scientists from October, and an increase in state spending on science to 4 per cent of all budget expenditures from next January. They have also asked for their representatives to be included in committees that are preparing the reforms.

Trade ruling could force India to reform patents

[NEW DELHI] In a decision with profound political implications for the Indian government, the disputes settlement board of the World Trade Organization (WTO) in Geneva ruled last week that patent protection in India is inadequate. The ruling is a victory

for the United States, which took India before the board last year for failing to honour its obligations under the Trade Related Intellectual Property Rights agreement (see *Nature* 383, 656; 1996).

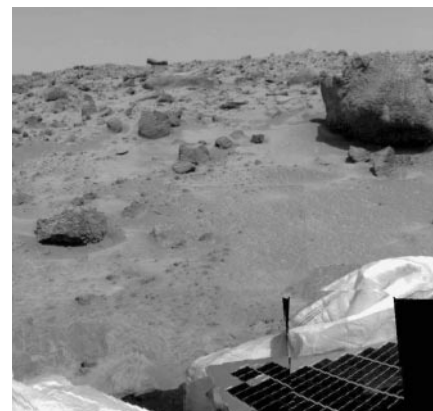
India's defeat means that WTO members can take retaliatory action that could adversely affect India's exports. The ruling also means that the government faces the task of having to pilot a bill through parliament to amend the Patent Act of 1970. The bill introduced by the previous government, which was blocked by the upper house, had lapsed when parliament was dissolved before last year's general election.

Soros funds US-Russian diagnostic projects

[MOSCOW] George Soros, the Hungarian-born philanthropist, has announced a donation of US\$3 million to the New York Institute of Public Health for a joint project with three Russian scientific organizations. The goal of the project is to set up diagnostic centres for rapidly and reliably locating the agents of infectious diseases — mainly tuberculosis — using up-to-date methods such as PCR techniques for 'reading' the bacterial genome.

Soros will provide an extra \$2 million next year if the project's managers find \$5 million from other sources.

Mars robot reveals signs of catastrophic flooding



[WASHINGTON] A foot-long, windblown boulder nicknamed "Barnacle Bill" (foreground, left of centre) was the first rock examined by Mars Pathfinder's roving robot, Sojourner, following its landing in the Ares Vallis region of Mars last Friday (4 July). Pathfinder's first images showed hilly terrain and a wide variety of rock types. Scientists said on Monday that they also see clear signs of past catastrophic flooding in the rounded shapes of some boulders and the angles at which they have been deposited. The rover was expected to spend at least a week taking stereo images and chemically analyzing rocks and soil within 100 metres of the landing site.

AP/NASA JET PROPULSION LABORATORY

Is magnetic fusion heading for ignition or meltdown?

Physicists have long been keen to harness the energy of thermonuclear fusion. The next planned step is a major international project to build an experimental reactor, ITER. But the project has recently been running into a quagmire of technical, economic and political uncertainties that will require all the efforts of its supporters to overcome.

For 50 years, researchers have sought to improve their understanding of how man could capture the power of fusion, the energy source of the Sun and all other stars. Their quest reaches a historic turning point next year, when the leading industrialized countries are due to decide if, when and where to construct the International Thermonuclear Experimental Reactor (ITER), a US\$10 billion testbed for the physics and engineering of magnetic confinement fusion.

For most of the past half-century, the favoured approach among fusion researchers has been the magnetic confinement of a fast-flowing loop of hot, ionized gas — or plasma — circulating in an evacuated, doughnut-shaped chamber called a tokamak. But there are many alternative approaches, and serious doubts persist about the feasibility of the tokamak concept which ITER will embody.

After billions of dollars of international investment in tokamaks, the best machines that have been built produce, at most, one-third of the energy that they consume. ITER is intended to break new ground not only by generating more power than it consumes, but by addressing some of the immense engineering challenges of operating a tokamak on a continuous basis.

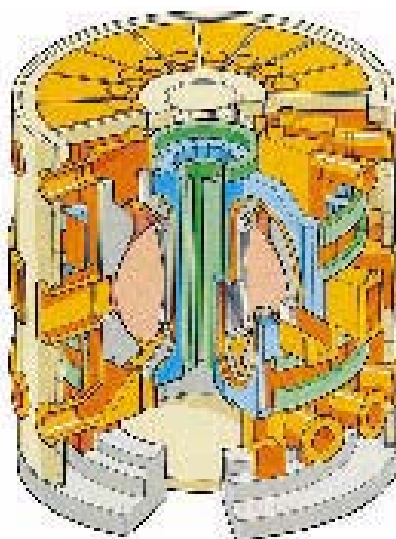
The reactor vessel of such a tokamak would have to survive intense neutron irradiation from the burning plasma, while transferring heat out of the plasma at an unprecedented rate. When its structure is eventually destroyed by the neutrons, it would have to be replaced on a modular basis, by remote control.

Some scientists and engineers from outside the fusion field see such problems as insurmountable. They point to the debilitating effect of a constant flux of highly energetic (14 MeV) neutrons on the tokamak chamber, and to the geometrical difficulties of transferring heat rapidly from the chamber to a coolant which would surround but not penetrate the reactor vessel.

Fusion advocates have long sought the construction of a special, high-energy neutron source to test the materials that a working fusion plant would require. But, they say, even without such a facility, materials researchers have been able to establish that vanadium and some steel alloys hold promise for use in such an environment.

"A number of the results have been very promising," says Bob Conn, dean of engineering at the University of California at San Diego. "When someone says it would be impossible, I don't think they know the data." He also believes that the heat transfer rate expected to be required at the edge of a tokamak — around 1 MW per square metre — is achievable. "Many systems today operate at many times that," he says.

Just as difficult to contend with are the environmental and licensing problems that



Vital statistics: the design of the proposed ITER tokamak envisages a reactor measuring 16 metres across, with a plasma current of 24 MA, fusion power of 1.5 GW, toroidal field of 5.7 T and a burn time of 1,000 seconds.

would surround the construction of a fusion-based power plant. Although it would produce no high-level nuclear waste, a fusion plant would produce intermediate and low-level waste, particularly when decommissioned, and it would incorporate new and untested technologies. The construction of such a plant might well encounter fierce public opposition.

But fusion advocates say that such issues, however daunting, must be tackled if man is to gain access to an energy source that could meet the needs of a future global population of 10 billion. "One-hundred-and-fifty years ago, equally insurmountable difficulties could have been identified when contemplating heavier-than-air flight,"

one correspondent wrote to *Physics Today* in May. "One is glad now that Osborne Reynolds, Ludwig Prandtl, N. Joukowski and the Wright brothers didn't give up."

When they started the engineering design assessment phase of the ITER project in 1992, the four partners — Europe, Japan, Russia and the United States — agreed that they would make a firm decision about its construction by July 1998. But, as governments face the need to control public spending, it has become apparent that they will not be ready to start building then. Through an informal group called the ITER Explorers, the partners are now looking at how to delay construction without draining the programme of momentum.

Budget cuts bite

Since 1995, when the Congress cut the budget for magnetic fusion research from \$360 million to \$240 million, it has been clear that the United States would not play a major role in building the experimental reactor. Last year, the best hopes for construction in Europe evaporated (see page 119). But the really bad news for the project came last month, when budget pressures in Japan led to a pledge that the Treasury would fund no major new scientific projects for three years (see page 116).

Serious international collaboration on magnetic fusion began at an Atoms for Peace conference in Geneva in 1958, when physicists at secret nuclear weapons laboratories in the United States and the Soviet Union exchanged notes on the peaceful use of fusion power. After the Second World War, as both countries raced to develop the fusion-based hydrogen bomb, their weapons scientists had sought means of harnessing the energy released when either the two nuclei of the hydrogen isotope, deuterium, or one atom of deuterium and one of tritium, are made to fuse together.

The key to the problem is the successful

On other pages | Japan budget shock: p116 | Rivals to the tokamak: p118 | Doubts in Europe: p119

confinement of deuterium or deuterium-tritium plasma while it is heated to 100 million °C or more — temperatures at which any physical containment material would vaporize. Scientists have pursued two approaches to this: magnetic confinement fusion, which relies on magnets to confine the plasma, and inertial confinement fusion, which compresses and heats a small amount of fuel *in situ*, giving it nowhere to go but inwards on itself.

The hydrogen bomb is an uncontrolled inertial confinement fusion device, compressed by a fission-based nuclear explosion. Weapons scientists have worked for more than 30 years to build a controlled device, to help them with their weapons simulations. In the United States and France, major efforts are under way to do this with powerful laser beams. But such lasers take hours to cool down after one 'shot', and an inertial fusion machine for power generation would more probably be driven by an accelerated beam of heavy ions.

Magnetic attraction

Magnetic confinement fusion, in contrast, had little military value, but apparent potential as the core of a continuous power generation process — if it could be achieved and sustained. Various schemes were proposed to contain plasma magnetically, and in the late 1960s Soviet physicists persuaded their colleagues in the West that their concept of a tokamak was the most promising one.

Scientific collaboration continued as experimental tokamaks were built around the world, for example at Princeton University in the United States, at the Joint European Torus (JET) at Culham in the United Kingdom, and JT-60 in Japan. In these, scientists gradually raised the temperature of the plasma and the amount of fusion energy produced. The collaboration culminated in the launch of the ITER project at the first summit meeting between the then US President Ronald Reagan and Mikhail Gorbachev, in Geneva in 1985.

Europe and Japan joined the project at once, allowing it to absorb their own respective plans to build larger tokamaks, while political and economic turmoil in the Soviet Union curtailed its financial contribution (although scientific involvement from Russia remains robust today). Unable to agree on a location for a single central laboratory, ITER divided its workload between three 'home teams' at Garching in Germany, San Diego in California, and Naka in Japan.

Partners follow different paths

The outcome of the project's six-year engineering design assessment has inevitably become a compromise between the interests of different partners, reflecting differing views of what the machine should aim to achieve. Broadly, the United States had (and

still has) a science-orientated fusion programme, and wanted a scientific experiment to enable the study of plasma at states known as 'burn' and 'ignition'.

In fusion of deuterium and tritium, four-fifths of the energy is released as neutrons and one-fifth is available as alpha particles, to reheat the plasma. If a tokamak's fusion 'gain' — heat out divided by heat in — reaches a value of five, the alpha heating will match the heat in, and the plasma is said to 'burn'. At ignition, the burn becomes self-sustaining, and no heat input is needed, so the gain is infinite. In a working fusion reaction, some energy would always be fed in to drive and control the plasma, and the gain would be about 25.

The Japanese and Europeans had more immediate worries about energy supply than the Americans, and a greater inclination to spend public money on large technology

development projects. They have been seeking a testbed for important fusion technologies, including new materials and efficient, superconducting magnets.

Many US critics continue to contend that ITER is the wrong machine. Two years ago, the President's Council of Advisors on Science and Technology (PCAST) argued that the partners should reconsider the project and try to build something smaller (see *Nature* 377, 567; 1995). The other partners rejected this proposal out of hand. "I'm confident that the machine we propose strikes the right balance between risk and meeting our objectives," says ITER's director Robert Aymar.

The criticism might have been contained if the US fusion programme had enough resources to pursue new experiments other than ITER. But it does not: magnetic fusion research has been under siege in the United

Japanese industry fights budget setback

Japanese hopes of hosting the ITER reactor received a severe setback last month from a budget reform plan that puts a ban on inviting the project to the country during the next three years (see *Nature* 387, 643; 1997). Proponents of the project in Japan still want the reactor to be sited there, but the reform plan has triggered public debate about the country's broader commitment to fusion research.

The budget reform has long been in the making, and underlines the fact that 'big science' projects face growing opposition from the powerful finance ministry. It follows a policy document on financial reform, published last year, which advocated a shift from large-scale — and often inefficient — public research and development efforts towards more flexible project funding, with a focus on research in universities.

The ITER project is said to have been raised during discussions at the ministry's advisory committee on public finance at the end of last year. This committee's recommendations are said to have been influential in setting the pace for the budgetary reform advocated by prime minister Ryutaro Hashimoto last month.

Despite the budgetary uncertainties, officials at the Science and Technology Agency (STA) who oversee ITER-related activities in Japan say the project is "inevitable" for the country's long-term energy prospects, and an "important opportunity" for Japan to immerse itself in large-scale international research collaboration. Past Japanese efforts in international cooperation in research, including the Human Frontier Science Projects, have only been partially successful.

The fact that the United States has

considerably reduced its commitment to the experimental reactor has enhanced Japan's interest in the project, for those who feel that it may reduce the risk of the project becoming entangled in political tensions. But Japanese officials admit that they have only limited experience of cooperating with Europe.

The STA, which recently came under severe public criticism for its handling of the Monju fast-breeder reactor accident (see *Nature* 386, 209; 1997), has another reason for promoting ITER. Almost all Japanese research on tokamak fusion has been done by the Japan Atomic Energy Research Institute, a national research centre overseen by the STA and employing more than 1,600 researchers.

But over the last two decades, the institute has come under increasing pressure from more powerful organizations — such as the Power Reactor and Nuclear Fuel Development Corporation, which is now responsible for most of Japan's nuclear energy research — as well as the private sector. According to some observers, the institute is in considerable need of an overhaul and a new lead project such as ITER.

The proposed fusion reactor has influential supporters in Japanese industry. Hitachi, Toshiba and Mitsubishi Electric — all prime contractors to the nuclear industry — have been developing expertise in fusion-related technologies since the early 1980s, and they expect to win important engineering contracts if the project is built in Japan.

Last February, Keidanren, an organization made up of the chairmen of major Japanese industrial groups, set up a Forum for the Promotion of ITER to lobby

States. It has been attacked by budget hawks, advocates of renewable energy, proponents of continued reliance on fission energy — and by the physics community itself.

The black art of plasma physics — which is rather like the famously difficult study of turbulent flow, only tougher — has been treated with suspicion by many mainstream physicists. “It is very empirical, and you can’t always do it from first principles. That causes a problem with other physicists,” says Anne Davies, head of fusion research at the US Department of Energy (DoE). She talks of a “chasm” between fusion science and the rest of physics.

But it is the hype that surrounded fusion, particularly in the aftermath of the 1973 energy crisis, that has alienated many scientists. “There is ill-feeling from many other fields of study,” says Richard Hazeltine, director of the Institute for Fusion Studies at

the University of Texas at Austin. He says that the fusion community has only itself to blame: “It is the result of us not being careful enough with our language when we advertise fusion. We should have been more careful in the things that we said.”

Doubts over design

Even within plasma physics, ITER is not universally popular. Bruno Coppi, a plasma physicist at the Massachusetts Institute of Technology, has been predicting for years that the machine will not ignite, and advocating an alternative design — the Ignitor — as a cheap and compact route to ignition.

But, as the US magnetic fusion budget has shrunk, there has been little money to develop alternatives to the tokamak. Competition between ITER and proposed US facilities, such as the now-abandoned Tokamak

Physics Experiment, has also taken its toll on US support for the international project.

Last year, new doubts about the technical integrity of the design surfaced. A physics-based model of plasma turbulence, developed by William Dorland and Michael Kotschenreuther of the Institute of Fusion Studies at Austin, suggested that ITER would not ignite (see *Science* **274**, 1600; 1996). This assessment damaged ITER politically, according to several fusion scientists, but it helped plasma physics by spurring the investigation and application of such models. Dorland and Kotschenreuther are now working in the ITER project although, as Dorland notes, “this is not to say that they accept our results”.

In fact, preliminary work with other supercomputer-based models of hot plasma is being cited by ITER advocates as refuting the research by Dorland and Kotschenreuther. According to Marshall Rosenbluth, a former member of the plasma physics department at the University of California at San Diego and now with the ITER Joint Central Team, this work suggests that the turbulence effects predicted by the Texas team will reach a saturation point, reducing by a factor of eight the rate of heat loss from the plasma that they predicted.

Where the Texan research models the plasma as a fluid, the new models simulate the kinetics of up to 100 million plasma ions directly, on a particle-by-particle basis. These models have been developed by Andris Dimits at Lawrence Livermore National Laboratory in California and Richard Sydora at the University of California at Los Angeles.

Plasma problems pinpointed

The DoE has established a regular, informal forum to discuss and attempt to resolve these issues. An April assessment of the current ITER design by the Fusion Energy Sciences Advisory Committee (FESAC), which advises the department on its fusion programme, said that uncertainties over plasma physics made it difficult to judge whether the reactor will confine plasma well enough to meet its design goals.

Conn, dean of engineering at the University of California at San Diego and lead author of the study, believes that the machine should still be built. “You can’t guarantee anything — that’s why it’s an experiment,” he says. The study said that the reactor’s fusion gain (Q) could fall anywhere between four — a serious disappointment — and infinity. “We have good confidence that Q will reach a value greater than 10,” says Conn.

In the United States — where the likely foreign support for the project has sometimes been overestimated — the news that Japan, as well as Europe, are seeking to delay construction has been received badly. Until recently, for example, Hazeltine says he was

the government to bring the reactor to Japan.

Until recently, criticism of ITER was largely restricted to local anti-nuclear groups and citizens’ movements in areas such as Tomakomai and Hokkaido, which are widely discussed as possible sites for the reactor. But influential voices are now joining the critics, including the country’s largest newspaper, *Yomiuri Shimbun*.

Last December, in an attempt to provide a broad political platform to discuss options in regard to ITER, the Atomic Energy Council (AEC) set up another ITER Forum, bringing together the main stakeholders in the project and the representatives of various social groups. A report of its discussions is due to be published later this year.

These discussions are expected to provide guidance and advice on the practicalities of and conditions attached to a Japanese initiative to host the reactor. But a political decision on whether the project will be invited to Japan may still be several years away. As the AEC has no budgetary authority, financial questions will not be addressed by the forum, even though budgetary constraints are likely to be a key issue in the decision on whether to bid to host ITER.

Evaluations of the reactor’s detailed design report by US, European and Japanese experts all suggest that present cost estimates may be over-optimistic. Critics in Japan say that the high construction costs in the country’s nuclear industry may push total costs even higher, perhaps doubling the present most widely quoted estimate in Japan of ¥840 billion (US\$7.3 billion).

Officials at the STA dismiss such estimates. But they admit that the cost issue

needs to be urgently readdressed on a site-specific basis.

“We need realistic cost estimates and a site-specific safety evaluation to judge whether we can express our intention to invite ITER,” says Satoshi Tanaka, director of the office of fusion energy at STA.

But some fusion scientists in Japan, while supporting the project in principle, argue that the reactor’s design is simply too expensive. “An overall design review with an eye to cost efficiency is a precondition for Japan to host ITER,” says Atsuo Iiyoshi, director of the National Institute of Fusion Science (NIFS) and president of the Japanese Society of Plasma Science and Nuclear Fusion Research.

Others have been openly attacking ITER, arguing that Japan should take a more balanced approach to the development of future energy sources. Given the technological difficulties surrounding nuclear fusion, they argue that the parallel development of various design options would be a better strategy.

As an example of this approach, Japanese universities have already used education ministry funding to build an impressive series of helical fusion devices. In this design the plasma stream is confined by a system of helical coils, an approach that has been abandoned in most countries except Japan and Germany.

The world’s largest helical fusion device, the Large Helical Device — with a magnetic field of 30,000 Gauss — is being completed at NIFS. But the budget for helical fusion has been only a fraction of what has been spent on tokamak fusion.

Robert Triendl





Aymar: tokamak still the best research bet.

confident that the experimental reactor would be built, one day. Asked whether he still believes this, he says: "No, I do not. I knew there were reservations in the United States but, when I learned that Europe and Japan were ready to delay construction, I took that to be an ominous indicator."

But supporters of the project continue to talk up the prospects. "I do believe that ITER will go forward," says the DoE's Davies. John Sheffield, chair of FESAC and a physicist at Oak Ridge National Laboratory in Tennessee, points out that a comparable delay preceded the construction of the JET by the member states of the then European Economic Community, and could be "quite helpful" in finalising the design.

But many other observers, inside and outside the programme, doubt that ITER could survive such a hiatus. In the United States, money could be hard to find for continued collaboration on design work if the

construction is subject to doubt. In recent informal discussions, the United States has insisted that any extension to the engineering design assessment should end after two years unless there is a firm commitment to construction. "We insist that this will not be open-ended," says Davies.

Open-ended or not, the extension is set to reinforce a growing split between the partners over their approaches to the project. The United States, which has said that it will not host the reactor, will use the delay to complete research which slowed down during the engineering design assessment, as the US contribution to ITER fell from \$80 million in 1995 to \$55 million this year. Europe and Japan are expected to use the extension to do site-specific design and to negotiate with licensing authorities about possible sites.

But, even as construction is delayed, none of the partners wants to pull the plug on an international collaboration that has probably run as long and as deep as that in any science discipline. Indeed, the scientific collaboration itself is a major reason why political supporters of the project, such as George Brown, the senior Democrat on the Science Committee of the House of Representatives, stand by it. "I think it will be built," he says.

Rivals that might oust the tokamak

If ITER is built, it will be the culmination of a 30-year period in which the Soviet idea of a torus-shaped, current-carrying ring of plasma, called a tokamak, has been the leading candidate to harness fusion power. If ITER collapses, alternative devices are poised to re-emerge as rivals to the tokamak.

Germany and Japan continue modest development programmes for the oldest such device, the stellarator. Sometimes known as a helical-fusion device, or heliotron, the stellarator looks and operates like a tokamak. But, where the tokamak relies on a current running through the plasma itself to induce the desired magnetic field, the stellarator uses an external, helical coil to induce the field.

In the United States, research on stellarators and other ideas for magnetic fusion — including the Elmo Bumpy Torus developed at Oak Ridge National Laboratory in Tennessee — has been squeezed out over the past decade as the fusion programme has shrunk from its peak of \$470 million in 1984 (worth more than \$700 million in today's money) to \$240 million today.

The only such idea still being pursued is the \$25 million National Spherical Torus Experiment, a tokamak tightly compressed into the shape of a sphere, being built at Princeton Plasma Physics Laboratory in New Jersey (see *Nature* 386, 313; 1997).

Some scientists advocate a compact but powerful tokamak aimed at the single goal of

achieving ignition. Bruno Coppi of the Massachusetts Institute of Technology continues to seek government support in the United States and in Europe for such an experiment, which he calls the Ignitor.

But the most politically viable alternative to the tokamak is the entirely distinct approach of inertial confinement fusion (ICF), in which a tiny pellet of deuterium and tritium fuel is compressed and heated to produce a small, controlled thermonuclear explosion. The US nuclear weapons programme is about to spend \$1.2 billion to build the National Ignition Facility (NIF) at Lawrence Livermore National Laboratory in California, which will use the world's largest laser beams to do just that.

But the lasers will take hours to cool after one 'shot', and any ICF-based power plant would be more likely to rely on an accelerated beam of ions to heat the target. The US Department of Energy non-weapons fusion programme spends \$7 million a year studying such ion drivers. The scientists involved want \$100 million to build an ion accelerator to develop the idea. They say that such a facility, plus success at NIF, could open the way for the development of inertial confinement fusion as an energy source.

In an ICF power plant, they imagine, the targets would drop freely in a waterfall of liquid metal coolant, which would seamlessly absorb heat, neutrons and blast as they were 'shot' by the accelerator. **C.M.**

But Brown's confidence will have to spread to others if the sums are to add up. Partners used to assume that construction costs would be split under a so-called "n+1" formula, with costs divided evenly but the host paying one extra share. When there were four partners, the split would be 40-20-20-20; when it became apparent that Russia was out, the split became 50-25-25. Now that the United States has backed off, all bets are off.

According to Davies, Japan has said informally that it will only pay half the costs as host, dashing earlier hopes that it would pay up to 70 per cent. The United States might contribute 5 per cent and the Europeans, according to Davies, have indicated that if the reactor is built in Japan they will not pay more than the United States. So how can it be built? "They might find some other partners," suggests Davies.

ITER weighed in the balance

President Bill Clinton has recently asked his council of advisors on science for what is expected to be a landmark review of strategic US energy research. Physicist John Holdren of Harvard University is leading the study, which is due to be completed in October.

The nuclear component of the review is likely to weigh ITER against other options in magnetic fusion research. It may also compare magnetic fusion's prospects with those of inertial confinement fusion — which no impartial body has attempted to do, in the United States, since key elements of inertial confinement fusion were declassified in 1990.

The fate of ITER and of magnetic confinement fusion research rests on how badly the industrial nations feel that they need a new, non-carbon emitting energy source. Confronted with low energy prices and public complacency over the issue, particularly in the United States, supporters of fusion are reduced to hoping that global warming will revive interest in strategic energy research.

Forty years after the public debut of magnetic fusion, its advocates have dropped their claims that it constitutes an easy route to a boundless energy supply. Today, their claims are more modest. "For the tokamak, I'd like to say what Churchill said about democracy," says Aymar. "It may look bad, but there is no other line of enquiry that would be better."

Asked if ITER will proceed, Sheffield expresses the mainstream US view: "It's obviously up to Europe and Japan," he says.

Balking at ITER's cost, the United States has passed the buck to the rest of the world. But like the US Congress, which is considering a modest restoration of fusion spending after recent sharp cuts, the rest of the world seems content to continue with fusion research as a kind of insurance policy — without taking the plunge and agreeing to invest \$10 billion in building the experimental reactor.

Colin Macilwain

European passions cool as bill increases

Long-standing scientific enthusiasm for fusion as a potential source of energy within Europe's close-knit fusion community has lately been on a collision course with a steady decline in the political appetite for paying the hefty bill for the ITER project.

In the past 30 years, Europe has spent ECU8 billion (US\$9 billion) on magnetic-fusion research. Scientists believe that they are close to reaping the rewards of this investment — provided that support is not withdrawn. "We are close to building a machine that can produce fusion," says Martin Keilhacker, director of the Joint European Torus (JET) near Oxford in the United Kingdom. "And we could have a prototype commercial machine in 50 years." But politicians remain sceptical, recalling similar claims made over the years that have failed to materialize.

The European Commission's fusion programme is often held up as a shining example of how European research collaboration can work effectively. It is a tightly coordinated effort that integrates research funded by national governments and the European Commission. The commission pays nearly half of Europe's ECU500 million annual expenditure on fusion research, administered through its atomic agency Euratom.

Hopes pinned on ITER

In the mid-1980s, Europe abandoned plans for its own experimental facility, the Next European Torus (NET), in favour of ITER, a larger version of the machine that was being considered for NET. ITER has now become the focal point of the entire programme. Other fusion technologies, such as inertial confinement, are only supported on a 'keep-in-touch' basis, and occupy a small fraction of the fusion budget.

So fusion scientists know that, if ITER were to disappear over the horizon, the fusion programme could well collapse. The most alarming sign that this could happen occurred last summer, when hopes that Europe might provide a host site for the ITER reactor diminished sharply after France and Germany jointly — and unexpectedly — announced that they would not offer candidate sites. The potential costs to the host had become too high (see *Nature* **382**, 288; 1996). France and Germany have the largest domestic research programmes in Europe.

A few months later, Italy, which has the third largest fusion research programme, offered to step into the breach. The Italian research minister, Luigi Berlinguer, suggested that Italy should use its underspent structural funds — the European Union's subsidies to poorer regions — to help develop potential sites in the south of the country.

A committee has since selected potential sites that meet the physical requirements of



Hotting up: the JET tokamak (above) has helped blaze the trail towards the ITER reactor.

the experimental reactor's current design. One candidate is said to be near Latina, south of Rome, and there are at least two others in the Puglia region on the Adriatic coast.

But Italian hopes may founder on the issue of money. Italy has not made effective use of the current round of structural funds earmarked by the government for research, which are worth hundreds of millions of dollars. But these cannot be redirected to a potential reactor site, as all contracts must be issued before the end of 1999.

And Italy cannot rely on the next round of structural funds being as generous, since their future is under debate between EU member states. Without support from structural funds, Italy would have to find its own money to support a site — and there is no reason to believe it would find this any easier than have Germany or France.

Canadian option

Canada, which participates in ITER through the EU partnership, has two promising sites. But the federal government is not ready to support them. It ended its financial support for Canada's fusion research programme earlier this year. Instead, proposals for the sites — by Lake Huron and Lake Ontario — are being promoted independently by the electricity utility Ontario Hydro.

Even if Europe cannot offer a site, it should remain a strong partner in the ITER programme, according to a recent independent report on fusion research, commissioned by the European Commission in Brussels and coordinated by Sergio Barabaschi, a former industrial researcher

and currently scientific adviser to the Italian research minister.

But to do so, the report calculates, the commission's fusion programme would need an increase in funding of 10 per cent above its current level of around ECU225 million a year. All signs point against such a significant rise in the next few years, although a recent, as yet informal, agreement between ITER's four partners to delay a decision about construction for two or three years means that the need for an increase could be delayed (see *Nature* **387**, 746; 1997).

This still leaves the question of whether Europe will be prepared to maintain even its current spending on fusion. The European Commission pays for 45 per cent of the European fusion programme from its five-year Framework research programmes. The rest is provided from national research budgets.

The content and budget of the next, fifth, Framework programme, which is due to be launched in 1999, are now being discussed. The commission is responding to a paper prepared by the Netherlands, holder of the union's rotating presidency for the first half of this year, outlining five options for fusion research. These range from complete abandonment of the ITER project to the building of the reactor without further delay.

The commission favours pressing ahead with the project, and supports delaying decisions about construction while continuing work on the development and licensing of potential sites (see *Nature* **387**, 746; 1997). It foresees extending funding for JET, which is currently supposed to close in 1999, to allow further studies there on plasma physics that would help optimize ITER operations.

Scientists are confident that this option would mean no significant disruption to the research and development programme in the near future. The commission hopes that, since the plan has no major budgetary implications, it will win the day.

But the union's two decision-making bodies, the Council of Ministers which represents member governments, and the European Parliament, could yet cause trouble. While most members of the parliament support fusion, some influential members have said publicly that the programme should be sharply cut, and several governments privately agree.

Researchers are watching the debate nervously, hoping that holders of the purse-strings do not keep their eyes closed to the progress that has been made. Keilhacker says: "When JET started [in 1983], the triple product [temperature, energy confinement time and plasma density] was 100-fold away from the minimum value predicted to achieve ignition. Now we are only four to five times short of the minimum value." **Alison Abbott**

Mine waste pollutes Mediterranean

Sir — Garcia-Guinea and Huascar¹ discuss the catastrophic environmental impact of huge quantities of waste materials from polymetallic ore mining in South America. A similar situation — fortunately without such dramatic consequences so far — affects the southwest Mediterranean coastline.

Spain, Morocco, Algeria and Tunisia have a variety of volcanic-related mineral deposits (especially oxides, sulphides and sulphosalts of base and precious metals). Southeast Spain has been mined since 3000 BC, and has more than 120 abandoned mines — the highest number in Europe — concentrated along approximately 300 km of the Mediterranean coastline. Immense quantities of metal sulphides (lead, zinc and silver in Cartagena, Sierra Almagrera), precious metals (gold and silver in Rodalquilar, Las Herrerías) and mercury–antimony (Valle del Azogue) have been extracted; some mines are still active (for example, the Las Herrerías open pit).

Over the past six years, I have led several mineralogical, geochemical and metallogenetic research projects in this region, in cooperation with the Spanish Comisión Interministerial de Ciencia y Tecnología and IUGS/Unesco^{2,3}. There are no regional environmental studies of the intense mining activity of the area. (There are more than 2,000 mining pits and 230 tailing dams and dumps in the province of Almería alone⁴.) There are also industrial and ore-dressing plants close to the coast.

Despite this huge volume of mining

waste, no geochemical mapping of the toxic element anomalies has been carried out — unlike in other major mining districts (for example the Iberian ‘Pyrite Belt’⁵, the Tees River basin⁶) — and no data exist about the geographical distribution patterns of the mining-related chemical elements. Monitoring studies of metal distribution are only local and are restricted to areas where the pollution has reached extremely high levels. The most critical case is Portman Bay (Spain): this is the most contaminated bay in the entire Mediterranean, and a perfect example of ecotoxic pollution of a coastal environment by mine tailings. The waste from mining operations was discharged directly into the inner part of the bay for more than 30 years, polluting the sea for a radius of several kilometres⁷.

On a larger scale, the southwest Mediterranean area is sadly contaminated, as has been shown by the high concentration of arsenic found in water, soils and sludges⁸, lead in bivalve shells (up to 100 parts per million, more than 30 times the average mean)⁹, methylmercury, selenium, cadmium, zinc and lead in dolphins¹⁰, lead and zinc in fish and other marine organisms¹¹ and so on. This pollution is also conditioned by the interaction of cyclone and anticyclone streams through the Straits of Gibraltar and from the north of Corsica. These streams converge in the Algerian–Provençal marine basin, meeting in the Alboran sea (where high quantities of cadmium and manganese

have been detected¹²), and cause an accumulation of mining tailings and industrial waste.

The most recent Unesco Science World Report stresses that environmental preservation and solving problems created by human activity is one of the most important challenges now facing mankind, and the World Conservation Union considers the Mediterranean a protected area. It is therefore of the utmost importance that the European Union and North African countries affected should monitor this environmental threat which is progressively poisoning rivers, aquifers, soils, beaches and the sea.

Jesús Martínez-Frías

Consejo Superior de
Investigaciones Científicas,
Museo Nacional de Ciencias Naturales,
Departamento de Geología,
c/ José Gutiérrez Abascal 2,
28006 Madrid, Spain
e-mail: mcnmf53@fresno.csic.es

1. Garcia-Guinea, J. & Huascar, M. *Nature* **387**, 118 (1997).
2. Martínez-Frías, J. et al. *Econ. Geol. Bull. Soc.* **87**, 444–446 (1992).
3. Martínez-Frías, J. et al. *Neues Jb. Miner. Abh.* **4–97**, 175–184 (1997).
4. ENADIMSA Dir. Gen. Ind. Ener. y Minas, 1–2, 645 (1986).
5. VanGeen, A. et al. *Geology* **25**, 291–294 (1997).
6. Hudson Edwards, K. et al. *Sci. Tot. Environ.* **194**, 437–445 (1997).
7. CEDEX http://almeria.cedex.es/CEPYC/bibl/memo/1994/memo_e.html
8. Navarro, M. et al. *Bull. Environ. Cont. Toxic.* **50**, 356–362 (1993).
9. Auernheimer, C. et al. *Arch. Sci.* **49**, 87–98 (1996).
10. Cardellicchio, N. *Wat. Sci. Technol.* **32**, 331–340 (1995).
11. Pastor, A. et al. *Mar. Pollut. Bull.* **28**, 50–53 (1994).
12. Statham, P. J. et al. *Nature* **313**, 565–567 (1985).

French research needs

Sir — Declan Butler describes the possible impact on French research of the change of government (*Nature* **387**, 645; 1997). He emphasizes that the Socialists have a more favourable attitude towards academic research than the previous government, but also that mandatory reforms would face “widespread resistance”, particularly if the system of life tenure were to be threatened.

This is probably a realistic view of the reaction of most civil servants. Budgets have been under pressure for a long time, and there have been no resources to enable postdoc researchers to build dynamic research teams. Foreigners are welcome in the Centre National de la Recherche Scientifique (CNRS) and other national organizations, which proves that it is not a chauvinistic system. It is, however, a rigid system, and there is no real alternative between getting a permanent position and

having to rely on one-year research and teaching posts in universities. Most new French PhDs have to leave French laboratories when they have completed their three years.

This rigidity reflects a general problem in the French economy, where there is no real tradition of short-term employment contracts (perhaps renewed several times), the most efficient way to manage projects in a fast-changing world. A worker must by law be offered a full-time post after two short-term contracts lasting a maximum of 18 months. Only a few semi-private institutions such as the Institut Pasteur are free to hire contract researchers of any nationality at will. Academic laboratories of the CNRS and others that would like to recruit postdoctoral students have to rely on a few 18-month positions offered by the Direction de la Recherche et des Etudes Techniques of the defence ministry, or on specifically designed short-term contracts reserved for foreigners (*postes rouges* or

roses). Most postdocs are foreigners with grants from foreign or supranational organizations.

This situation has its origins in the French *grandes écoles* system, which leaves French industry in the hands of well trained engineers who usually lack contact with fundamental research (except in large companies), and the academic system in the hands of scientists with no incentives or financial means to transform their research results into new products and jobs.

It is to be hoped that a consensus on these basic facts, which imply the need for a different type of funding for academic research, will be accepted by both sides of the political spectrum.

Xavier Michalet

Laboratoire de Biophysique de l'ADN,
Institut Pasteur,
25 Rue du Dr Roux,
75724 Paris cedex 15, France
e-mail: michalet@pasteur.fr

Just follow the acid chain

Gottfried Schatz

Mitochondria have to import almost all of the proteins they need for their function as the cell's energy producers. A combination of biochemistry and genetics now reveals a tiny but crucial cogwheel in the import machinery.

A typical eukaryotic cell contains roughly 10,000 different proteins. About ten per cent of these are in the mitochondria — the double-walled organelles in which respiration occurs. Because mitochondria synthesize only about a dozen proteins, they must import 99 per cent of their proteins from the cytoplasm. But how do they recognize these proteins, and how are they transported across the two membrane barriers? On page 195 of this issue, Dietmeier *et al.*¹ describe a yeast mitochondrial protein that seems to work in both processes and thereby links them. This protein, which they call Tom5, is an integral protein of the mitochondrial outer membrane, and it is so small that it would be missed by most commonly used analytical methods.

There is, however, nothing small about what Tom5 does. It seems to be an important cogwheel in the dauntingly complex mitochondrial protein-import machine², which works as follows. A protein that is destined to be imported to the interior of a mitochondrion is usually synthesized as a precursor, having a transient amino-terminal targeting signal that can form a strongly basic amphiphilic helix. As the precursor emerges from the ribosome, it binds to one of the cytosolic chaperone proteins, and it is then transferred to proteins that act as 'protein-import receptors' at the mitochondrial outer membrane. These receptors deliver the precursor, amino terminus first, to a protein-conducting channel in the outer membrane. Once the amino-terminal targeting signal has negotiated this barrier, it inserts into a protein-conducting channel in the inner membrane. This presumably occurs in response to the pull of the electric potential (negative inside) across that membrane. The precursor chain — which at this stage may span both membranes — is then pulled completely into the matrix space by an ATP-driven translocation motor that is attached to the inner side of the inner-membrane transport channel.

Mitochondria therefore have two separate protein-transport systems — the 'Tom' system in the outer membrane (for transport across the outer membrane), and the 'Tim' system in the inner membrane (Fig. 1). The Tom system consists of a hetero-oligomeric

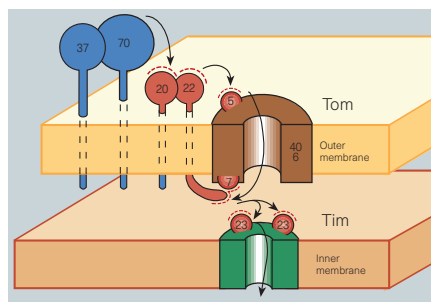


Figure 1 Dietmeier *et al.*¹ have identified Tom5, a small protein that plays a crucial part in a 'binding relay' (black arrows) for the import of proteins into mitochondria. The Tom system in the outer mitochondrial membrane consists of four receptors (Tom37, Tom70, Tom20 and Tom22) and a transmembrane protein channel composed of Tom40, Tom6, Tom7 and probably also Tom5. The basic targeting signal of some precursor proteins binds to Tom37/70 and then, through electrostatic interactions, to the 'acid bristle' receptors Tom20 and Tom22. (Other precursors bind directly to Tom20/22.) The receptor-bound precursor is then transferred to Tom5 and, after passage through the Tom channel, to several different binding sites on the inner face of the outer membrane. These sites include an internal Tom22 domain and possibly Tom7, as well as one or more unidentified sites. Transport to the Tim system may be facilitated by binding of the targeting signal to an acidic domain of Tim23. Acidic domains constituting the binding relay ('acid chain') are in red and are bordered by minus signs.

Tom channel, and at least four receptor proteins. Although the Tom and Tim systems are not firmly linked to each other, they interact dynamically to form a molecular pontoon bridge. This helps a targeting signal emerging from the Tom channel to enter the Tim channel, through which it is then moved by an electric potential. But the outer membrane lacks a transmembrane potential. So how is a receptor-bound precursor threaded into and across the Tom channel? Enter Tom5.

Dietmeier *et al.*¹ found that when the outer membrane is solubilized with the detergent digitonin, Tom40 — the core component of the outer-membrane protein-transport channel — co-fractionates with

three small proteins termed Tom5, Tom6 and Tom7. Their earlier work had already indicated that Tom6 and Tom7 are subunits of the Tom channel, which modulate its protein conductance³. But what about Tom5?

Microsequencing of the protein (M_r 5,980) revealed it to be the product of a previously unidentified nuclear yeast gene. Inspection of the amino-acid sequence, along with biochemical studies, led the authors to conclude that Tom5 is an integral outer-membrane protein with a carboxy-terminal transmembrane anchor and an acidic amino-terminal region that is exposed to the cytosol. When they deleted the *TOM5* gene, yeast cells became temperature-sensitive for growth. Moreover, isolated mitochondria that lacked Tom5 imported precursor proteins with greatly reduced efficiency, even though they were normal in other respects. Clearly, a hitherto undiscovered component of the mitochondrial protein-import machine had been found.

Where in this machine does Tom5 work? To answer this question, Dietmeier *et al.* first used all the tricks of the trade to check which of the previously characterized translocation intermediates were no longer formed efficiently in mitochondria lacking Tom5, or in wild-type mitochondria that had been treated with an antibody to Tom5. The Tom5-deficient mitochondria bound precursor proteins as efficiently as wild-type mitochondria, but they transferred bound precursors into the import channel of the outer membrane only very slowly. A similar message came from chemical crosslinking experiments — Tom5 was most effectively crosslinked to an early import intermediate that was still exposed on the outer membrane and had not yet inserted its amino terminus across the inner membrane. These results identified Tom5 as a link between the known receptors and the outer-membrane import channel.

Tom5 could be regarded as yet another import receptor on the mitochondrial surface, but it differs from the four known receptors in at least two ways: it acts after them, and it seems to be firmly associated with Tom40. Tom40 is efficiently co-immunoprecipitated by an antibody against Tom5, and Tom5 protects an amino-terminal region of Tom40 from proteolytic attack. So Tom5 may be an acidic subunit on the outer mouth of the Tom channel that accepts precursors with their basic targeting signals from the acidic receptors Tom20 and Tom22. In this way, Tom5 could facilitate the insertion of the signal into the channel. Being both a receptor and a subunit of the Tom channel, Tom5 may be part of a 'binding relay' of acidic binding sites. This relay could capture the precursors that are destined for import and feed them into the protein-transport channel of the outer membrane.

But the binding relay mechanism does

not seem to end there. The basic targeting signals can also bind to the inner face of the outer membrane⁴, and the authors cite their unpublished data which confirm that part of this binding is to an acidic Tom22 domain that protrudes into the intermembrane space⁵. Tom7 may contribute yet another *trans*-binding site. So, as sketched by the black arrows in Fig. 1, protein transport across the outer membrane may be driven by binding of the targeting signal to a series of strategically placed acidic binding sites of increasing avidity. A binding relay mechanism could also mediate transfer of the precursor to the inner membrane — the awaiting outer mouth of the Tim channel has exposed acidic domains on Tim23, which bind targeting signals⁶. Even the cytosol may be part of this proposed binding array, as the acidic cytosolic protein Mft2 (whose 'acid bristles' resemble those of Tom20) controls the import of some artificial mitochondrial precursors *in vivo*, and binds mitochondrial targeting signals *in vitro*⁷.

The binding relay model begs two questions. First, why do precursors not short-circuit this array by binding immediately to Tom5? Some short-circuiting may, in fact, be tolerated, because yeast cells lacking either Tom5 or Tom20 can still grow slowly at room

temperature. In other words, they can probably still import proteins into their mitochondria. But massive short-circuit on the mitochondrial surface may be prevented by the cytosolic chaperone to which the precursor is bound⁸. Ordered binding to the intramitochondrial sites might be safeguarded by the topological arrangement of these sites. Second, what clears the bound precursor from the last member in the array? Logical candidates are the potential across the inner membrane and the ATP-driven protein-import motor on the matrix side of the inner membrane. Viewed in this light, Tom5 is the newest and smallest link in an 'acid chain' that guides basic mitochondrial-targeting signals to their final destination. □

Gottfried Schatz is in the Biozentrum der Universität Basel, Klingbergstrasse 70, CH-4056 Basel, Switzerland.

1. Dietmeier, K. *et al.* *Nature* **388**, 195–200 (1997).
2. Schatz, G. *J. Biol. Chem.* **271**, 31763–31766 (1996).
3. Hönlinger, A. *et al.* *EMBO J.* **15**, 2125–2137 (1996).
4. Mayer, A., Neupert, W. & Lill, R. *Cell* **80**, 127–137 (1995).
5. Bolliger, L., Junne, T., Schatz, G. & Lithgow, T. *EMBO J.* **14**, 6318–6326 (1995).
6. Bauer, M. F., Sirrenberg, C., Neupert, W. & Brunner, M. *Cell* **87**, 33–41 (1996).
7. Cartwright, P., Beilharz, T., Hansen, P., Garrett, J. & Lithgow, T. *J. Biol. Chem.* **272**, 5320–5325 (1997).
8. Komiya, T., Rospert, S., Schatz, G. & Mihara, K. *EMBO J.* (in the press).

strain resulting from fault slip, following years of slow strain accumulation. Elastic rebound remains a central tenet of modern theories of earthquake generation.

The triangulation measurements central to Reid's analysis remain invaluable, because advances in our understanding of earthquake mechanics and vastly improved computational capabilities allow use of the data in modelling of the earthquake in ways that were not imaginable in his day. Thatcher and his colleagues at the US Geological Survey have been at the forefront of this research, and have now refined¹ earlier modelling of the slip that occurred on the San Andreas on 18 April 1906.

This study adds previously unanalysed data at the northern end of the 1906 rupture near Cape Mendocino (Fig. 1). The San Andreas is located offshore between Point Arena and Shelter Cove, and the interpretation of the fault geometry near Shelter Cove has been controversial⁵. Assuming the main trace of the fault runs through Shelter Cove, the triangulation data suggest 8.6 m of fault slip in this region, the maximum slip anywhere along the 1906 rupture.

But it is data from the southeastern end of the rupture that yield the most significant results, for this is the location of the magnitude-6.9 1989 Loma Prieta earthquake. In this region it is possible to compare the strain released in the 1906 earthquake with strain accumulation in the years between 1906 and 1989. Geological and geodetic measurements suggest that the San Andreas fault in this region slips 19 ± 4 mm yr⁻¹ when averaged over many earthquake cycles⁶. Mapping indicated 1.5 m of slip in this area during the 1906 event, leading to some forecasts of a high probability of a moderate earthquake on this segment of the fault^{7–9}.

The reasoning was quite simple: if the San Andreas slips 19 mm yr⁻¹ on average, and if the 1906 slip in this region was 1.5 m, then the strain released by the 1906 earthquake should have recovered by 1985. However, there is, of course, considerable uncertainty associated with both the long-term slip-rate and the 1906 co-seismic slip. Furthermore, our understanding of earthquake physics does not require the earthquake to occur when the earthquake strain-drop recovers. The so-called time-predictable model¹⁰ is nonetheless widely used in long-term forecasting, and deserves careful testing. It is for this reason that the slip in 1906 near Loma Prieta is so important.

Although the surface mapping indicated 1.5 m of slip, the fault trace in this area is complex and it is possible that additional slip was distributed over a fault zone a few hundred metres wide, and was thus missed. Here is where the triangulation measurements are crucial. These data show that the Loma Prieta moved 1.2 m to the southeast in 1906, consistent with 2.5 ± 0.4 m of slip on a

Seismology

New insights into old earthquakes

Paul Segall

Geodetic survey measurements conducted in the late nineteenth and early twentieth centuries continue to improve our understanding of the great San Francisco earthquake of 18 April 1906. These data are crucial to understanding the seismic strain history of the greater San Francisco Bay region, as exemplified in the latest work from Wayne Thatcher and colleagues, published in the *Journal of Geophysical Research*¹.

The 1906 earthquake was a milestone in the history of seismology. To many Californians it is the archetypal Big One, although amongst twentieth-century earthquakes it is not even close to being the biggest — the 1960 Chilean earthquake, the largest instrumentally recorded earthquake, had a seismic moment 350 times greater. Nevertheless, the 1906 earthquake was critical for what it revealed about the physics of earthquakes. Geologists quickly recognized that over 300 km of the San Andreas fault had slipped up to 6 m during the quake² (see map, Fig. 1). This observation helped to establish that fault slip caused earthquakes, rather than being a manifestation of secondary ground failure.

The transient seismic waves unleashed by the slip in 1906 devastated San Francisco. Permanent displacements of the Earth's crust, which remained after the shaking

stopped, caused other, albeit less dramatic, problems. For example, roads were beheaded by the fault. Permanent ground displacement distorted precise survey networks which had been painstakingly measured to map the coastline of California and San Francisco Bay. These surveys, begun in the 1850s, determined the relative latitude and longitude of landmarks with a precision of a few parts in a million (a few decimetres over a distance of 100 km). With earthquake displacements of many metres occurring in the quake, it was 'decided to repair the old triangulation, damaged by the earthquake, by doing new triangulation'³.

The 1906 epicentral region was resurveyed by triangulation in 1906 and 1907 by the US Coast and Geodetic Survey. When the data were analysed they revealed that fault-parallel motions extended tens of kilometres from the San Andreas. The horizontal gradient in these displacements showed a decrease in shear strain acting on the fault. Furthermore, the measurements made between the 1850s and 1880s showed that shear strain had been building up on the fault in the decades before the earthquake. From these data, Harry Fielding Reid formulated the elastic rebound theory⁴, in which earthquakes are caused by sudden release of shear



Figure 1 The greater San Francisco Bay region, showing the course of the San Andreas fault and the main locations discussed in the text. The 1906 earthquake was a consequence of rapid slip along more than 300 km of the San Andreas in this part of California.

vertical fault extending to a depth of 10 km (ref. 11). Thatcher *et al.*¹ investigated a range of plausible fault geometries and show that slip at depth along the Loma Prieta segment must have been between 2.3 and 3.1 m, much more than the 1.5 m of observed surface slip. According to the simple time-predictable model, the next earthquake was not due until well into the next century. It has been suggested that large uncertainties in the geodetic displacements prevent accurate determination of fault slip¹². The computed displacements, however, have larger uncertainties than the estimated fault slip, because the fault model implies additional constraints on the deformation pattern.

The issue of whether or not the Loma Prieta earthquake was anticipated⁷⁻⁹ is further complicated by the observations that it occurred on a dipping fault that many believe to be distinct from the San Andreas fault, and that it had roughly equal amounts of vertical and horizontal slip. Neither of these features was predicted. The previous large earthquake in the region, which occurred in 1865, appears to be distinct from both the 1906 and 1989 events. From seismic intensities¹³ and geodetic displacements¹⁴, it seems that the 1865 earthquake may have occurred on a thrust fault northeast of the San Andreas.

We are left with two messages. First, although the elastic rebound theory implies a basis for long-term earthquake forecasting, the best available data indicate that the history of earthquakes along the southernmost stretch of the 1906 rupture has been more complicated than Reid would have predict-

ed. Second, not only are data collected nearly 150 years ago essential in deciphering the seismic history of the region, but we should look back on the nineteenth-century surveyors with admiration for their thoroughness in archiving and documenting their data. □

Paul Segall is in the Department of Geophysics, Stanford University, Stanford, California 94305, USA.

1. Thatcher, W., Marshall, G. & Lisowski, M. *J. Geophys. Res.* **102**, 5353–5367 (1997).
2. Lawson, A. C. (ed.) *The California Earthquake of April 18, 1906* Vol. 1 (Rep. State Earthquake Investigation Commission, Carnegie Inst. Washington, Washington DC, 1908).
3. Hayford, J. F. & Baldwin, A. L. in *The California Earthquake of April 18, 1906* Vol. 1 (ed. Lawson, A. C.) 114–115 (Rep. State Earthquake Investigation Commission, Carnegie Inst. Washington, Washington DC, 1908).
4. Reid, H. F. in *The California Earthquake of April 18, 1906* Vol. 2 (ed. Lawson, A. C.) 16–28 (Rep. State Earthquake Investigation Commission, Carnegie Inst. Washington, Washington DC, 1910).
5. Brown, R. D. *Bull. Seismol. Soc. Am.* **85**, 100–110 (1995).
6. Working Group on California Earthquake Probabilities *Probabilities of Large Earthquakes in the San Francisco Bay Region, California* (Circ. 1053, US Geol. Surv., Reston, 1990).
7. Lindh, A. *Preliminary Assessment of Long-Term Probabilities for Large Earthquakes Along Selected Fault Segments of the San Andreas Fault System in California* (Open File Rep. 83-63, US Geol. Surv., Reston, 1983).
8. Sykes, L. R. & Nishenko, S. P. *J. Geophys. Res.* **89**, 5791–5800 (1984).
9. Scholz, C. H. *Geophys. Res. Lett.* **12**, 717–719 (1985).
10. Shimazaki, K. & Nakata, T. *Geophys. Res. Lett.* **7**, 279–282 (1980).
11. Segall, P. & Lisowski, M. *Science* **250**, 1241–1244 (1990).
12. Sykes, L. *Proc. Natl Acad. Sci. USA* **93**, 3732–3739 (1996).
13. Tuttle, M. P. & Sykes, L. R. *Bull. Seismol. Soc. Am.* **82**, 1802–1820 (1992).
14. Yu, E. & Segall, P. *J. Geophys. Res.* **101**, 16101–16118 (1996).

Apoptosis

Placing death under control

David Wallach

Unlike the gloomy connotation of the death of an organism, the death of an individual cell is an integral and continuing part of normal physiology¹. It is also a major form of defence, as sometimes the only way for the immune system to eradicate pathogens is to sacrifice the infected cell. When the timing of cell death is inappropriate, however, havoc may ensue. So understanding the mechanisms that regulate cell death is just as important as understanding those responsible for the actual killing. A newly identified protein, described on page 190 of this issue by Irmeler *et al.*², and in *Immunity* by Shu *et al.*³, appears to act as a regulator of one of the principal ways by which immune mechanisms cause cell death — namely its induction by receptors belonging to the tumour necrosis factor (TNF)-receptor family^{4,5}. This is a highly active research area, as exemplified by the fact that three other groups⁶⁻⁸, including my own, are shortly to publish on the same protein.

Death is a rather uncommon consequence of receptor triggering; usually, the process is concerned with the transmission of life-enhancing signals such as those that stimulate cell growth and division. Nonetheless, the emerging knowledge of induction of cell death by TNF receptors indicates that it happens in the same way as all other receptor-induced effects — that is, through a series of protein–protein bindings. The first involves the binding of specific ligands to the extracellular domains of the receptors. This is followed by sequential binding of cytoplasmic proteins to the intracellular domains of the receptors, leading eventually to activation of enzymatic function in some of these proteins. In the case of death induction, the activated enzymes include caspases, a family of cysteine proteases whose members occur in cells as latent precursors, becoming

activated early in the process of programmed cell death and being central to its development⁹.

The intracellular protein interactions triggered by the death-inducing receptors can be attributed to two structural motifs within the proteins concerned. Both motifs are able to associate with homologous regions in other proteins, and thus prompt binding of such proteins to one another (reviewed in ref. 10). One motif, the ‘death domain’ (DD), is found in several death-inducing receptors of the TNF family, including CD95 (also called Fas or Apo-1), CD120a (the p55 TNF receptor) and others. It also occurs in several cytoplasmic adaptor proteins that bind through this domain both to receptors and to each other (as, for example, in the MORT-1/FADD and TRADD adaptors).

The other motif, the ‘death effector domain’ (DED), is found in MORT-1/FADD upstream of the DD; it also occurs in duplicate in two caspases, caspase 8 (MACH/FLICE/Mch-5) and caspase 10 (Mch-4/FLICE-2). Binding of these two caspases to MORT-1/FADD through association of their DED motifs, and consequent activation of the caspases by their proteolytic cleavage (apparently by self-processing), are thought to be critical steps in the initiation of the killing process.

This multiplicity of interacting proteins provides points of ramification in the signalling pathway, allowing induction of different effects by the same receptor. Thus, several receptors of the TNF family that induce cell death can also activate the transcription factor NF- κ B, thereby turning on genes whose products provide resistance to cytotoxicity by these receptors. This latter activation involves TRAF-2, as well as NIK, a protein kinase that binds to it. TRAF-2 also binds two proteins, c-IAP1 and c-IAP2,

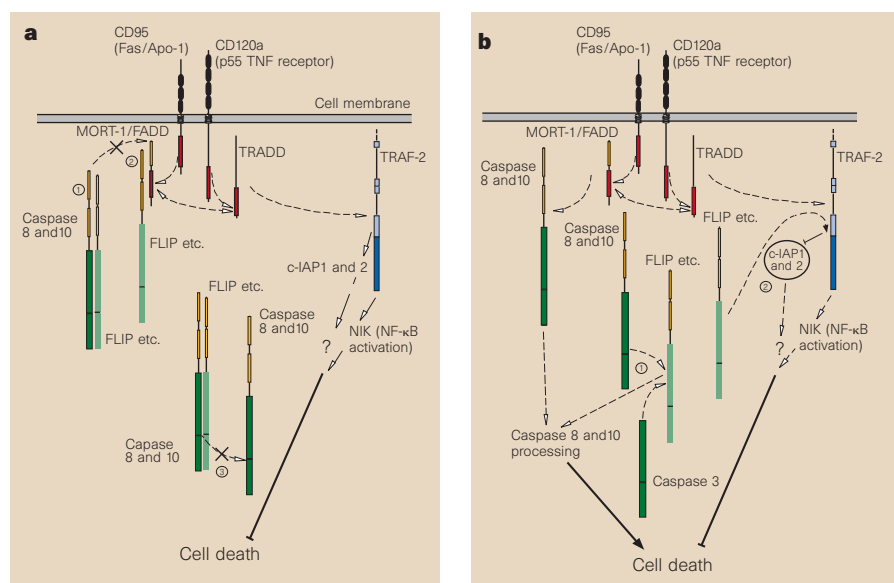


Figure 1 The newly discovered regulatory protein, variously called FLIP², Casper³, FLAME⁶, CASH⁷ and I-FLICE⁸ (here referred to as FLIP etc. for simplicity), may be involved in blocking the induction of cell death or in activating it. It acts downstream of receptors of the tumour necrosis factor (TNF) family such as CD95 and CD120a. **a**, Proposed mechanisms of blockage of cell-death induction. 1, inhibition of binding of caspases 8 and 10 to the adaptor protein MORT-1/FADD by the binding of FLIP etc. to any of these proteins; 2, inhibition of caspase 8 and 10 self-processing by the association of their protease-precursor regions with the corresponding region in FLIP etc. Receptors of the TNF family can also activate the transcription factor NF-κB, a process which involves the NIK protein kinase and TRAF-2, and is thought to confer protection against death. In addition, TRAF-2 binds the proteins c-IAP1 and c-IAP2, which may also provide protection against death by an unknown mechanism. **b**, Proposed mechanisms of cell-death activation. 1, triggering of caspase 8 and 10 processing by the simultaneous binding of FLIP etc. to these enzymes and to caspase 3; 2, displacement of c-IAP1 and c-IAP2 from TRAF-2.

which are thought to provide cells with resistance to death mechanisms.

In seeking additional regulatory proteins that specifically affect the initial event in death induction by members of the TNF-receptor family, Irmeler *et al.*² and Shu *et al.*³ concentrated on the terminal point of ramification in the pathway, namely, binding of caspase 8 and caspase 10 to the adaptor protein MORT-1/FADD. Their search led to independent identification of the same protein, which they respectively call FLIP and Casper. This protein contains DED motifs at its amino terminus and through them it can bind to other DED-containing proteins. The gene encoding the protein was localized to region 33–34 of the long arm of human chromosome 2, within 100 kilobases of the caspase 8 and caspase 10 genes (refs 2, 6 and my group's unpublished data).

The FLIP/Casper protein occurs in two sizes (splice variants). The shorter one is essentially composed of the DED motifs. The longer one contains, in addition, a region that closely resembles the carboxy-terminal protease-precursor regions in caspase 8 and caspase 10. In FLIP/Casper, this region lacks several of the sequence features required for protease activity. But it is able to bind directly to the protease-precursor region in caspase 10, as well as to caspase 3, another member of the caspase family that participates in the receptor-induced death process. Surprisingly, it can also

bind TRAF-2, displacing the putative death-regulatory proteins c-IAP1 and c-IAP2.

Expression of the long and short splice variants of FLIP/Casper, or of its mutants by introduction of their complementary DNA into cells, had pronounced but varying effects on the cell-death mechanisms activated by TNF receptors. In some cultured cells, it strongly inhibited cell death induction; in others, it triggered cell death.

Perhaps not surprisingly, then, the two groups have quite different views about the physiological role of FLIP/Casper (see Fig. 1). Irmeler *et al.* think that the death resulting in some cells from expression of the protein is artificial, and that the protein's sole function is to inhibit cell-death induction. They call it 'cellular FLIP', on the assumption that it acts like virus-produced inhibitors of cell-death induction, vFLIPs (viral FLICE-inhibitory proteins), which also consist of duplicate DED motifs^{11–13}. vFLIPs block death induction by interfering with the binding of DED-containing caspases to MORT-1/FADD, and Irmeler *et al.* suggest that FLIP/Casper works in the same way. Moreover, they say that the binding of the carboxy-terminal region in FLIP/Casper to caspase 8 and caspase 10 may constitute a further inhibitory mechanism, preventing the proteolytic self-processing of the caspases.

In contrast, Shu *et al.* propose that

NATURE

100 YEARS AGO

The chapter on resistance of cycles is very instructive. The conclusions are represented graphically, so that any one non-algebraically-minded can grasp the enormous importance of air resistance at high speeds. Extrapolating from these curves, it is seen that a man who can drive his machine under present conditions through the air at, say, 30 miles an hour, would, if road and machine resistance only had to be met, be able to drive at 330 miles an hour, or if he can actually go 20 miles an hour, he would be able to drive his machine 100 miles an hour. ... The writer would like to propose a method to enable great speeds to be attained, which, however, is of spurious interest, since in real cycling the wind resistance must be overcome. All that is necessary is that a large box or small house with glass sides big enough to entirely surround the rider, but with a safe margin, should be dragged by steam or other power along at gradually increasing speeds until the rider shows that he is beginning to lag. Of course, there would be no floor or bottom to the box, and it should be made so that it would clear the ground by any predetermined amount. It might be safer if the house had no back.

From *Nature* 8 July 1897.

50 YEARS AGO

Despite progress in education, the great majority of people seem to have but an inkling of the part played by chemistry and its sister sciences in improving the material conditions of their existence, and very few have any real interest in the subject; they will grasp for an aspirin, grab for a 'nylon' stocking or a coat of many colours, they make full use of the modern rapid means of communication, and relish the margarine that now masquerades as butter, without a thought that these innovations are the outcome of long and patient research in chemistry, physics or biology. One may excuse such ignorance and apathy in those who were born in Victorian days, but there is less excuse for the younger generations, although these also have been handicapped at school in their range of interests by too much concentration on the distant past.

From *Nature* 12 July 1947.

FLIP/Casper acts solely as a death activator, prompting activation of caspase 10 upon binding to it simultaneously with the binding to caspase 3, and perhaps also further enhancing death by displacement of the putative death-inhibiting c-IAPs from TRAF-2.

What of the work of the other three groups? Like Irmeler *et al.*, Alnemri and his colleagues⁶ observed an inhibitory effect of the protein on both CD95- and CD120a-induced cytotoxicity, and, similarly, suggest that it acts as an anti-apoptotic protein. They call it FLAME. In our own tests⁷, however, the protein (which we call CASH) exhibited both protective and marked cytotoxic effects, depending on the cell line examined. Finally, Dixit's group⁸ call the protein I-FLICE and see inhibition only.

Taken together, the data in the five reports indicate that the marked variations observed in the levels of FLIP/Casper/CASH/ FLAME/ I-FLICE in cells, depending on their states of differentiation and activation, constitute an important determinant of the effectiveness of cell killing by members of the TNF-receptor

family. But the markedly different interpretations of the various groups involved underscores a severe limitation in our ability to infer the physiological function of a newly discovered gene from its *in vitro* function. Clearly, further work is necessary to clarify not only what is happening *in vitro* but how these observations relate to the cell death that occurs in the body during health and disease. □

David Wallach is in the Department of Membrane Research and Biophysics, Weizmann Institute of Science, Rehovot 76100, Israel.

1. Jacobson, M. D., Weil, M. & Raff, M. C. *Cell* **88**, 347–354 (1997).
2. Irmeler, M. *et al.* *Nature* **388**, 190–195 (1997).
3. Shu, H.-B., Halpin, D. R. & Goeddel, D. V. *Immunity* **6**, 751–763 (1997).
4. Smith, C. A., Farrah, T. & Goodwin, R. G. *Cell* **76**, 959–962 (1994).
5. Nagata, S. *Cell* **88**, 355–365 (1997).
6. Srinivasula, S. M. *J. Biol. Chem.* (in the press).
7. Goltsev, Y. V. *et al.* *J. Biol. Chem.* (in the press).
8. Hu, S. *et al.* *J. Biol. Chem.* (in the press).
9. Henkart, P. A. *Immunity* **4**, 195–201 (1996).
10. Wallach, D. *Trends Biochem. Sci.* **22**, 107–109 (1997).
11. Thome, M. *et al.* *Nature* **386**, 517–521 (1997).
12. Bertin, J. *et al.* *Proc. Natl. Acad. Sci. USA* **94**, 1172–1176 (1997).
13. Hu, S., Vincenz, C., Buller, M. & Dixit, V. M. *J. Biol. Chem.* **272**, 9621–9624 (1997).

lie only in galaxies, there should be a close connection between the total amount of starlight in a system and the total mass of heavy elements⁴.

These objects are interesting not just in themselves, but for what they tell us about the Universe. The heavy elements probably can't escape the deep gravitational potential well of the cluster, so changes in the composition of the cluster gas with redshift (and thus with distance and look-back time) constrain the history of star formation in the Universe⁵. Also, clusters probably form in a 'bottom-up' fashion⁶, such that massive clusters form out of smaller clusters and galaxies. If the Universe is dense enough to be closed, massive clusters should form relatively late and so should evolve strongly with cosmic time.

Clusters of galaxies are strong gravitational lenses⁷, and their distortion of the images of background galaxies provides a measurement of the mass and mass distribution in the cluster. Another mass measurement can be derived from the X-ray temperature and density profile, assuming that the gas is in hydrostatic equilibrium. Both the X-ray and lensing data show that most of the mass in any cluster is some form of dark matter. For several well-studied clusters^{8,9} the optical light is a fairly accurate tracer of the dark matter, and therefore there is a strong connection between dark matter and the galaxy distribution. X-ray and gravitational-lens mass measurements of several nearby clusters, with redshifts $z < 0.4$, give mass-to-light ratios of $\sim 100\text{--}300h_{50}^{-1}$ (where the value for the Sun is 1, and h_{50} is the Hubble constant in units of $50\text{ km s}^{-1}\text{ Mpc}^{-1}$). This is only about a quarter of that required to close the Universe¹⁰ and so is used as an argument for an open Universe.

The discovery by Hattori *et al.* of a distant cluster ($z \sim 1$) with a very high mass-to-light ratio whose gas is considerably enriched in iron is a significant break with expectations, contrary to all these apparently established ideas.

The authors looked for X-ray emission from the region of a gravitationally lensed quasar, MG2016+112, using two X-ray satellites: ASCA, an X-ray astronomy spectroscopic mission; and Rosat, which has considerably better spatial resolution than ASCA. The large angular separation of this lensed quasar had led Narasimha *et al.*¹¹ to predict that the splitting was due to a massive gravitational lens, on the scale of a cluster mass. But sensitive optical searches showed no evidence for a cluster¹². So the discovery by Hattori *et al.* of a luminous X-ray-emitting cluster is unexpected.

In X-ray emission, this cluster is very similar to lower-redshift ($z < 0.4$) clusters, having similar size, luminosity, temperature and iron abundance. Using the X-ray data to obtain a mass via the equation of hydrostatic

X-ray astronomy

How one galaxy can be a cluster

Richard Mushotzky

On page 146 of this issue¹, Hattori *et al.* report the discovery of a distant cluster of galaxies that is massive and bright in X-rays, but surprisingly dark optically — containing only one visible galaxy, it has extremely anomalous optical properties (Fig. 1). In most theories of structure formation, massive clusters at high redshift should be rich in galaxies.

Clusters of galaxies are the largest gravitationally bound systems in the Universe. They

were originally found by searching for large numbers of galaxies in a small patch of sky, but they are also luminous X-ray sources. Most of the visible matter is in the hot X-ray-emitting gas between galaxies, rather than in the stars². In nearby clusters, the gas has about three to ten times the mass of the stars, and is enriched in heavy elements such as silicon and iron to roughly a quarter to a half of solar levels³. Because these elements are produced in stars, and stars are thought to

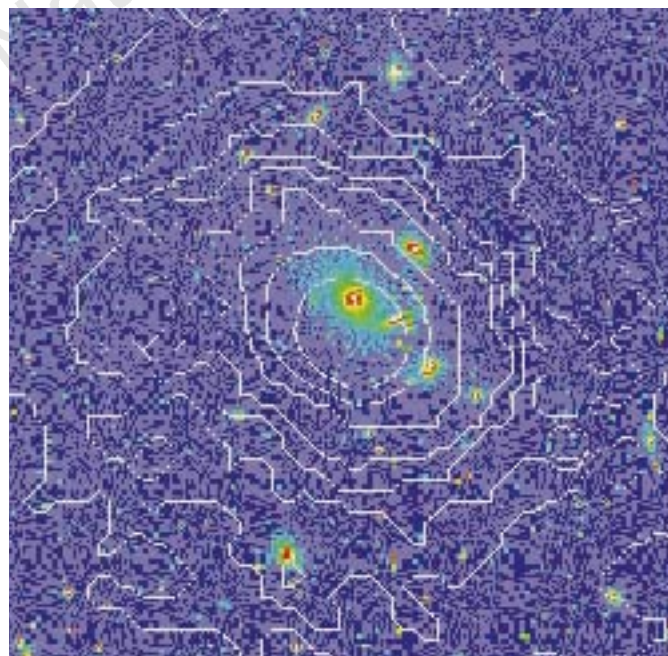


Figure 1 An empty cluster. In X-rays (countours), MG2016 appears as a massive galaxy cluster; but at optical wavelengths (false colours) only one obvious galaxy is present. It also has a high content of heavy elements, presumably produced by stars — but what stars?

equilibrium, Hattori *et al.* find good agreement with the gravitational lens model. The absence of many luminous galaxies thus gives a very large mass-to-light ratio of $\sim 3,000$ — as the authors say, a “dark cluster”. The mystery deepens when one realizes that the relatively large iron abundance, > 0.4 times that of the Sun, requires the existence of many stars, which are not seen now.

Optical and near-infrared images of the MG2016 field already reach down to the magnitude range of average galaxies¹³ at $z \sim 1$. So unless more sensitive observations find a rich cluster at magnitude limits just below the ones previously obtained, we must accept the existence of a massive dark-matter potential without luminous galaxies.

The high iron abundance in MG2016 indicates that most of its heavy elements (or ‘metals’) were formed at much higher redshifts. The lack of detected galaxies, combined with the redness of the one bright galaxy, shows that this cluster contains very few of the massive stars needed to produce the observed metals. Such stars have a lifetime of less than two billion years, and so the metals must have formed at a period of time corresponding to the observed age of the cluster plus the time for these massive stars to stop shining. This gives a predicted period of metal formation at $z \sim 2$, consistent with other indications from X-ray cluster metallicity measurements and recent deep galaxy

redshift surveys. But no one knows why the lower-mass stars, which should have been formed along with the high-mass stars, are absent in this cluster.

In conventional cosmological models, having a just-closed Universe dominated by cold dark matter, massive clusters should only be about 1/30th as common at $z \sim 1$ as at present. This new discovery does not give an indication of the true space density of these systems, but future work will, and in doing so will provide strong constraints on cosmological models. □

Richard Mushotzky is at NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

- Hattori, M. *et al.* *Nature* **388**, 146–148 (1997).
- Jones, C. & Forman, W. *Annu. Rev. Astron. Astrophys.* **20**, 547–585 (1982).
- Loewenstein, M. & Mushotzky, R. *Astrophys. J.* **466**, 695–703 (1996).
- Arnaud, M., Rothenflug, R., Boulade, O., Vigroux, R. & Vangioni-Flam, E. *Astron. Astrophys.* **254**, 49–64 (1992).
- Mushotzky, R. & Loewenstein, M. *Astrophys. J.* **481**, L63–L66 (1997).
- White, S. D. M. in *The Early Universe with the VLT* (ed. Bergeron, J.) (Springer, Berlin, 1997).
- Valdes, F., Jarvis, J., Mills, A. & Tyson, A. *Astrophys. J.* **281**, L59–L61 (1984).
- Tyson, A. & Fischer, P. *Astrophys. J.* **446**, 55–58 (1996).
- Squires, G. *et al.* *Astrophys. J.* **482**, 648–658 (1997).
- Carlberg, R., Yee, H. & Lingston, E. *Astrophys. J.* **478**, 462–475 (1997).
- Narasimha, D., Subramaniam, K. & Chitre, S. *Astrophys. J.* **315**, 434–439 (1987).
- Lawrence, C., Neugebauer, G. & Matthews, K. *Astron. J.* **105**, 17–21 (1993).
- Cowie, L., Songaila, A., Hu, E. & Cohen, J. *Astron. J.* **112**, 839–864 (1996).

Geochemistry

Neon signs mark Australian plume

Bernard Marty

Mantle plumes are among the most spectacular processes that generate magmas at the surface of the Earth. They are giant upwellings of hot material from deep in the mantle, probably from below the convective mantle region that drives plate tectonics¹. Plumes produce long volcanic chains, such as the Hawaii–Emperor chain; and they are thought to hit the lithosphere eventually, to form large continental flood-basalts, such as the Deccan and the Ethiopian traps, over just a million years or so. These plumes participate in some way in the building of the continental crust, yet for many continental volcanic provinces, the geochemistry does not provide clear evidence for the existence of a plume. Matsumoto *et al.*, on page 162 of this issue², describe a new geochemical tool that has allowed them to detect a mantle plume under Australia.

Why are they hard to detect? If plumes originate from a deep region, they should be clearly ‘primitive’ in their composition. Unfortunately, the lavas produced are not made of primitive material; instead they appear to come from mixing between the

upper mantle and recycled oceanic or continental crust¹. In addition, melts and fluids derived from a plume head are modified by interactions with the continental lithosphere and the underlying mantle, so initial compositions can be difficult to identify using classical geochemical tools.

Matsumoto *et al.*² instead use a newly explored isotope system, that of the rare gas neon, to identify the contribution of plume-derived fluids to recent volcanic rocks (erupted 25,000 years ago) in southeastern Australia. They find a tenuous plume signal in one mineral, apatite. Other, co-existing mantle-derived minerals contain volatiles from the shallower lithospheric mantle. Their observation implies that mantle volatiles of different origins co-exist on a sub-millimetre scale for long periods. The authors also suggest two stages of fluid/melt evolution, with the trapped plume signal coming from the earlier and deeper stage.

The isotopic compositions of the rare gases have already provided powerful insights into the structure of the mantle. These inert elements were trapped in the

interior of the Earth during accretion, and have since been escaping into the atmosphere through volcanoes. Several rare-gas isotopes are produced by nuclear reactions in the solid Earth, allowing quantification of the time-varying rates of rare-gas exchange between the mantle and the atmosphere.

The simplest isotope system is that of helium. The ratio between the primitive isotope ^3He , trapped during Earth accretion 4.5 billion years ago, and the radiogenic isotope ^4He , generated constantly by the decay of its long-lived radioactive parents uranium and thorium, is the same in all mid-ocean-ridge basalts (MORBs) worldwide. So the source of MORBs, presumably the upper mantle, must have been thoroughly homogenized through convection. The $^3\text{He}/^4\text{He}$ ratios of plume-derived volcanic minerals are much more variable, but are often higher than those of MORBs, suggesting that the source of plumes is a mantle region less degassed than the MORB source. Because mantle degassing is a function of magma generation and therefore of mantle convection, convection appears to be less vigorous in the plume-source region.

The three isotopes of neon are a more subtle and more powerful system than helium¹. Mantle neon is a mixture of three different components (Fig. 1): atmospheric, primitive mantle, and nucleogenic. The first two of these were trapped early in the Earth’s history in the atmosphere and the mantle, and are distinguished by their $^{20}\text{Ne}/^{22}\text{Ne}$ ratio, which for atmospheric neon is 9.8 and for neon trapped in the solid Earth ≥ 13 , possibly as high as the solar value of 13.8. The third, nucleogenic neon, is mostly ^{21}Ne , constantly produced by neutrons and α -particles (from uranium and thorium) reacting with ^{18}O and ^{24}Mg nuclei.

Different mixtures of any two ‘end-member’ sources lie on a straight line. The compositions of MORBs follow a well-defined mixing line (Fig. 1, overleaf), between that of air and that of the MORB mantle source³. Plume-derived minerals (as defined by Hawaiian lavas) have a steeper mixing line, indicating a plume source richer in primitive neon than the MORB source^{4,5}. So neon isotopes should be able to distinguish between the two.

Australia provides a test case. Age progression among volcanoes⁶, high heat flow, and areas of low seismic velocity at depths of 80–140 km, all suggest the presence of a mantle-plume track beneath southeastern Australia. Furthermore, the geochemistry of the ‘newer volcanics’ in southeastern Australia is consistent with a plume head⁷ being under the continent now. These lavas contain fragments of mineral assemblages from 30–60 km down in the continental lithosphere, much shallower than the source of plume material, which must be at least several hundred kilometres deep. Some of

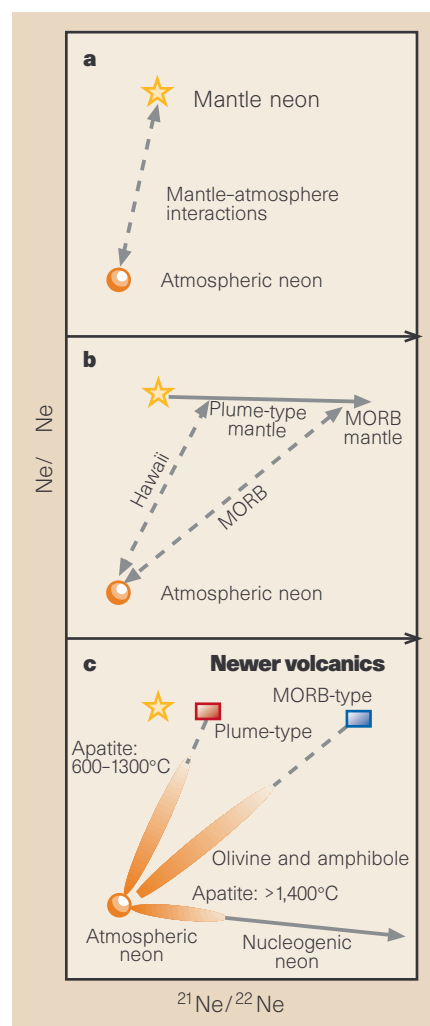


Figure 1 Three isotopes of neon point to different mantle origins. Mixtures between different end members fall along straight lines. **a**, In the early Earth, solar-composition neon was incorporated in the mantle, and neon with a lower $^{20}\text{Ne}/^{22}\text{Ne}$ ratio was left in the atmosphere. **b**, Nucleogenic ^{21}Ne has gradually moved the mantle composition towards the right-hand side of the diagram. The slopes of the mixing lines between the atmosphere and the mantle end members decrease with increasing loss of primitive ^{22}Ne from the mantle source, so the steeper slope of the correlation observed in Hawaiian lavas means that their mantle source is less degassed than that feeding MORBs. **c**, Young volcanic rocks from southeastern Australia contain apatite that seems to have the imprint of a plume component beneath the continent.

these fragments contain apatite and amphibole, which are thought to be products of metasomatic interaction, in which fluids in the uppermost (lithospheric) mantle alter the rocks.

Matsumoto *et al.* found MORB-type helium and neon compositions in olivines and amphiboles. The apatite, however, is profoundly different. The $^3\text{He}/^4\text{He}$ ratios are much lower than those of co-existing olivines and amphiboles, but this probably results

from production of ^4He in the apatite before eruption, since apatite has much higher concentrations of the radioactive parents uranium and thorium than have the co-existing minerals. Thus, any primordial mantle helium signature that may have existed is now masked by a recent radiogenic contribution.

Neon behaves more interestingly. Neon extracted from apatite at temperatures of more than 1,400 °C — probably the neon component residing in the silicate network — is a mixture of atmospheric and pure nucleogenic neon. With that assumption, the amount of accumulated ^{21}Ne would imply that the apatite formed about a million years before the lava erupted. At temperatures of 600 °C to 1,300 °C, however, the extracted neon lies on the Hawaiian correlation, characteristic of a plume-type component (Fig. 1). To explain the co-existence of two unequilibrium neon components in a single mineral, Matsumoto *et al.* propose that the plume component is concentrated in fluid inclusions, and is for this reason less subject to secondary radiogenic contributions.

Although this hypothesis seems reasonable, a number of points need clarification. First, optical investigation can decide whether the fluid inclusions were trapped when the apatite crystallized, or later. Second, the location of plume-type neon in fluid inclusions could be verified by crushing apatite under vacuum and analysing the released gases. Unfortunately, the authors could not perform this experiment due to the limited quantity of apatite available. Third, the low $^3\text{He}/^4\text{He}$ ratios were ascribed to a radiogenic contribution, yet trapping in fluid inclusions is supposed to reduce that. This discrepancy between helium and neon may be due to different mobilities of the two elements during heating or residence in the lithosphere. Again, a crushing experiment may help to solve this problem. Fourth, only apatite contains plume-type neon, whereas amphibole, which also results from metasomatic interaction, shares MORB-type gas signatures for both helium and neon.

This study nevertheless presents an interesting case: to identify giant processes taking place beneath our feet, it is sometimes useful to consider a very rare, but noble, geochemical tracer, in a well-hidden location. □

Bernard Marty is at the Centre de Recherches Pétrographiques et Géochimiques, 15 Rue Notre-Dame des Pauvres, BP 20, 54501 Vandoeuvre lès Nancy Cedex, France.

- Hofmann, A. W. *Nature* **385**, 219–229 (1997).
- Matsumoto, T., Honda, M., McDougall, I., Yatsevich, I. & O'Reilly, S. Y. *Nature* **388**, 162–164 (1997).
- Sarda, P., Staudacher, T. & Allegre, C. J. *Earth Planet. Sci. Lett.* **91**, 73–88 (1988).
- Honda, M., McDougall, I., Patterson, D. B., Dougeris, A. & Clague, D. A. *Nature* **349**, 149–151 (1991).
- Hiyagon, H. *et al.* *Geochim. Cosmochim. Acta* **56**, 1301–1316 (1985).
- Wellman, P. & McDougall, I. *Tectonophysics* **23**, 49–65 (1974).
- McDonough, W. F., McCulloch, M. T. & Sun, S. S. *Geochim. Cosmochim. Acta* **49**, 2051–2067 (1992).

Daedalus

Volatile hardware

In the early days of cosmic-ray research, rabbits in a normal wooden hutch were compared with rabbits kept down a mine in a lead-lined hutch, to shield them from the rays. The shielded rabbits died faster, suggesting that the rays were beneficial. It later turned out that they had somehow been poisoned by their lead shielding.

Children can likewise be poisoned by lead paint in their homes. One explanation is that they habitually pull flakes of paint off walls and windows, and eat them. But Daedalus recalls that in the nineteenth century, many people (including Napoleon) became poisoned by arsenical wallpaper. They didn't eat the paper: it was colonized by moulds which metabolized the arsenical pigment to volatile arsenic trimethyl. The hapless victims breathed in the liberated vapour. Now lead can also be methylated to a volatile compound — lead tetramethyl, an antiknock agent for petrol. So Daedalus reckons that rabbits and children alike owed their fate to lead tetramethyl evolved from the mouldy walls around them.

Musing on this theory, Daedalus began to dream of an ecologically virtuous petrol, enriched with natural, organic, mould-fermented lead tetramethyl. But he now has a better application. Tin also has a volatile tetramethyl derivative; solder is an alloy of lead and tin; and the world's vast turnover of rapidly obsolescent electronic gadgets is held together by soldered joints. So DREADCO's biologists are scraping moulds from decaying plumbers' shops, abandoned lead-chamber works and neglected tin-plate factories, in the search for strains that can methylate solder to a volatile mixture of methyl derivatives.

The implications are profound indeed. At present, our rubbish-disposal systems are burdened by a swelling flood of outmoded computers, sound systems, radios and TV sets. Daedalus will dump this whole stream in a special 'electronic compost heap' seeded with his solder-methylating mould. It will undo all the joints, and the electronic junk will fall apart. The resulting shower of clean resistors, capacitors, chips and circuit boards will be sized and sorted by shaker-screens, while the tetramethyl vapours will be condensed and demethylated to pristine solder. Everything will be recovered for reuse. The only snag is the disaster that would follow if the mould escaped, and began to chew its way relentlessly through the tin cans, lead batteries and soldered electronic gadgets on which the whole of our civilization is based.

David Jones

Obituary

Alfred Hershey (1908–97)

One of the founding fathers of molecular biology

The 1940s saw the origins of molecular biology and the beginnings of a proper understanding of living systems. Until then, biological phenomena were either reduced to questions of chemistry or treated in abstract. There seemed to be no middle ground between the solidities of biochemistry and the abstractions of genetics, and it was not customary to ask how cells knew what they were doing. So the great leap forward largely depended on scientists who were not wedded to the technologies of an established discipline, and could focus their attention on the simplest possible system — the bacterium *Escherichia coli* and its viruses. Delbrück, Luria and Hershey (a physicist, a physician and a microbiologist) played a crucial part in these developments, and for this they shared a Nobel prize in 1969. With the death of Al Hershey on 22 May, the last of the three has gone and an era has come to an end.

Hershey spent the first 16 years of his career as an instructor in bacteriology at Washington University, St Louis. The head of the department was Bronfenbrenner who, like many people in those days, believed that viruses were proteins that somehow triggered their host cell into making more of the same so were not in any sense alive. Bronfenbrenner's influence ensured that Hershey's early publications (on bacterial growth and virus–antibody interactions) were rather undistinguished. Certainly, they gave no hint of what was to come.

But in 1943 Hershey met Max Delbrück, just at the time when Delbrück and Salvador Luria were struggling with the mathematics of mutation in bacteria. Three years later, at the age of 38, Hershey reported that bacterial viruses contain genetic determinants which occasionally undergo spontaneous, inheritable changes (mutation). Shortly afterwards, he showed that these determinants display genetic linkage and form linear arrays — just like the chromosomal genes of higher organisms. This established that bacterial viruses are a simple version of all living creatures and, as Delbrück had imagined, that they could be used to investigate the nature of the gene.

In 1950 Hershey moved to the Carnegie Institute of Washington's unit at Cold Spring Harbor and he remained there for



the rest of his career. His studies on linkage and recombination in the genes of the T2 bacteriophage became increasingly complicated and, some years later, culminated in a review of almost impenetrable obscurity; privately, he admitted that the review was his way of saying that he could not understand what was going on. So, he changed to studying the chemical components of phage, and their fate during infection of a bacterium. Within a year, he and his assistant, Martha Chase, had shown that it is the DNA — not the protein — that carries the viral genes and is responsible for the transfer of information from one generation of particles to the next. This came to be known as the 'blender' experiment, because they had used an ordinary kitchen blender to separate the empty protein shells of the infecting virus particles from the infected bacteria.

Hershey's evidence that genes are made of DNA rather than protein was less precise than Oswald Avery's (for the bacterial transforming principle). However, Hershey's experiment identified the chemistry of a group of conventional genes, whereas Avery's result applied to a single obscure trait (the ability of a bacterium to make a particular surface polysaccharide). For that reason many people found Hershey's experiment more persuasive. But together, the two sets of results inaugurated the science of molecular biology. As Hershey wrote, the development of molecular genetics showed repeatedly that "some redundancy of

evidence was needed to be convincing" and "diversity of experimental materials was often crucial to discovery".

His next eight years were spent filling in the biochemical and genetic details of the life-history of bacteriophage T2. Finally, following the discovery by Pete Davison that DNA molecules are fragile in solution, Hershey set about developing a way to work out the size of native, unbroken molecules. It was a desperately difficult project, but his studies were the essential precursor for much easier work by others.

Al Hershey was, in many ways, the perfect experimentalist. No project was too difficult for him to tackle single-handedly, no thought too much effort. Right to the end of his career he carried out his own experiments — he would even spend hours washing up the laboratory glassware (because, as he said, it temporarily spared him the burden of thought). It was, perhaps, an extension of this principle that made him say that 'Hershey heaven' was to find an experiment that worked, and then to do it again and again.

He seemed to feel the responsibility of the world on his shoulders: only if he laboured to the utmost of his ability could everyone else avoid being misled. He divided every day in half; one half usually starting at 9 a.m. and the other at about 9 p.m., with dinner, sleep and breakfast interposed between each half. At almost any time of day or night, therefore, you might find him either sitting motionless in his chair, thinking and gazing at the ceiling, or cocooned in the privacy of some experiment or blissfully cleaning up afterwards. His powers of concentration tended to make him taciturn and absent-minded. But when he did talk, it was as if a searchlight were being trained on a dark landscape. He seemed to approach every subject, serious or trivial, as if he had never thought about it before (or, sometimes, even heard of it). This gave his utterances a penetrating originality, which often caused a gap in the conversation while his listeners desperately revised their opinions and reordered their thoughts.

His final project was perhaps the hardest. Various circumstances forced him to retire in his mid-60s, and make space for younger scientists. Typically, he made that final transition without a murmur and, for 20 years, sought satisfaction in other versions of Hershey heaven — the clearing of forest to make a vegetable garden, and a search for the underlying meaning of Western music.

John Cairns

John Cairns, formerly of the Harvard School of Public Health in Boston, is at Hollygrove House, Wilcote, Charlbury OX7 3EA, UK.

Latent pigments activated by heat

Organic pigments are used as colourants in paints, plastics, printing inks¹ and in the electronics industries². They can be distinguished from dyes in that they are insoluble. Whereas dyes readily dissolve to give homogeneous coloration, the application of organic pigments frequently requires a time- and energy-consuming dispersion step. Here we report a method for converting a pigment into a suitable precursor, a 'latent' pigment. Like a dye, the latent pigment is readily soluble or homogeneously dispersible in the application medium, and on subsequent thermal treatment it can be transformed *in situ* to the parent pigment.

The insolubility of organic pigments in the majority of solvents or matrices is due either to the existence of strong intermolecular π - π interactions in planar, closely packed systems (for example, the phthalocyanines), or to other strong intermolecular interactions, most commonly hydrogen bonding as in 3,6-diphenyl-1,4-diketo-pyrrolo[3,4-*c*]pyrrole^{3,4} (DPP; Fig. 1a). We have created⁵ latent pigments whose hydrogen bonding is transiently disrupted by the addition of protective groups. We use a protective group such as the *t*-butoxycarbonyl radical, because it breaks down the hydrogen-bonded pigment network to yield soluble derivatives and is easily removed by thermal treatment⁶.

Using DPP (compound 1), a well-known commercial red pigment, as a model compound, we succeeded in converting the pigment into *N,N'*-bis-(*t*-butoxycarbonyl)-3,6-diphenyl-1,4-diketo-pyrrolo[3,4-*c*]pyr-

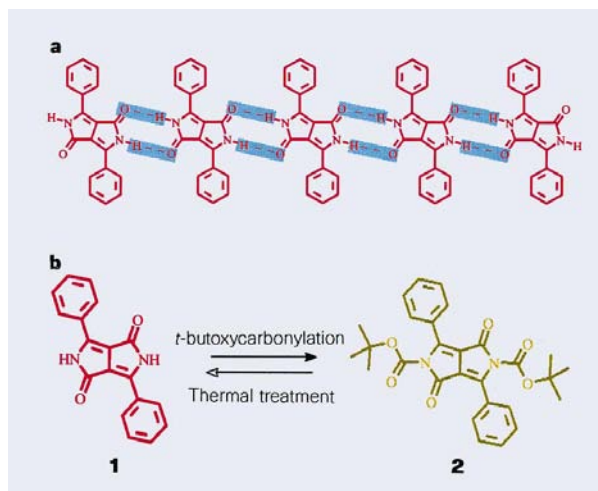


Figure 1a, Schematic illustration of the hydrogen-bonding network of DPP. **b**, Insoluble parent pigment (1) and soluble latent pigment (2).

role (2) in high yield under remarkably mild conditions, by simply stirring the pigment in tetrahydrofuran (THF) at room temperature with two equivalents of di-(*t*-butyl)-dicarbonate, using *N,N'*-dimethylaminopyridine as a catalyst^{5,7} (Fig. 1b). The carbamate formation in 2 is accompanied by a large hypsochromic shift of the maximum absorption in the solid state from 540 nm (ref. 4) to 439 nm.

The solubility of the latent pigment surpasses by several orders of magnitude the solubility of the parent pigment in common organic solvents. It has a solubility of 36 g per litre in xylene and 120 g per litre in cyclopentanone at room temperature, whereas 1 is practically insoluble ($<10^{-3}$ g per litre). Moreover, compound 2 undergoes simultaneous thermal elimination of both protective groups at about 180 °C (Fig. 2a) to afford the parent pigment in high purity and quantitative yield, with CO₂ and isobutene as gaseous products of the elimi-

nation. Analytical data and powder X-ray diffraction diagrams of DPP formed by the thermal decomposition of the soluble precursor are identical to those of DPP prepared by standard synthetic procedures³.

We prepared thin, yellow films (~1.5 μ m) of *p*-hydroxy-polystyrene containing roughly 40% of the latent pigment by spin-coating a cyclopentanone solution of both components on quartz substrates. Thermal treatment for 2 min at 180 °C resulted in the *in situ* regeneration of the parent pigment, as evidenced by optical spectroscopy (Fig. 2b), providing highly transparent red films. Further thermal treatment at 180 °C or higher temperatures (up to 220 °C) produced no change in the absorption spectrum. The optical spectrum of DPP formed by this process has a profile identical to that of a DPP thin film from vapour deposition⁴. Transmission electron micrographs of these thin films show the homogeneous distribution of the freshly formed pigment particles (<20 nm) throughout the whole matrix (Fig. 2c).

Our method combines the advantages of dye-like solubility with the solid-state properties of insoluble pigments¹, and could have significant advantages for pigment technology because it simplifies the application process but still allows homogeneous distribution of pigment nanoparticles in a substrate⁸. Our technique is applicable to most commercial high-performance organic pigments bearing functional groups⁹ capable of reacting with standard protective group reagents. In addition, our approach might become a powerful tool for probing and even influencing the role of hydrogen bonding in crystal engineering and molecular recognition¹⁰.

J. S. Zambounis, Z. Hao, A. Iqbal

Pigments Division, Ciba Specialty Chemicals Inc.,
CH-4002 Basel, Switzerland
e-mail: john_s.zambounis@chbs.mhs.ciba.com

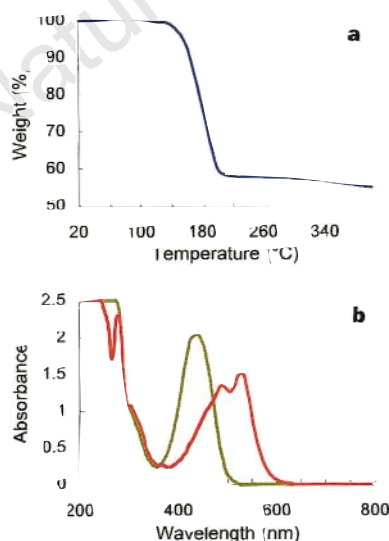


Figure 2a, Weight loss of *N,N'*-bis-(*t*-butoxycarbonyl)-DPP as a function of temperature (rate, 10 °C min⁻¹; onset, 163.6 °C; midpoint, 177.7 °C; end point 191.8 °C; step height, 40.7%). **b**, Ultraviolet-visible absorption spectra of a 1.5- μ m-thick *p*-hydroxy-polystyrene film containing 40% *N,N'*-bis-(*t*-butoxycarbonyl)-DPP, spin-coated (yellow) and after 2 min at 180 °C (red). **c**, Transmission electron micrograph of an ultramicrotome-cut cross-section of a 1.5- μ m-thick *p*-hydroxy-polystyrene film containing 40% *N,N'*-bis-(*t*-butoxycarbonyl)-DPP, after 2 min treatment at 180 °C.

- Herbst, W. & Hunger, K. *Industrial Organic Pigments* (VCH, Weinheim, 1993).
- Gregory, P. *High-Technology Applications of Organic Colorants* (Plenum, New York, 1991).
- Iqbal, A. *et al. Bull. Soc. Chim. Belg.* **97**, 615–643 (1988).
- Mizuguchi, J. & Rihs, G. *Ber. Bunsenges. Phys. Chem.* **96**, 597–606 (1992).
- Zambounis, J. S., Hao, Z. & Iqbal, A. US Patent No. 5,484,943 (1996).
- Rawal, V. H. & Cava, M. P. *Tetrahedr. Lett.* **26**, 6141–6142 (1985).
- Grehn, L. & Ragnarsson, U. *Angew. Chem. Int. Edn Engl.* **23**, 296–301 (1984).
- Schädeli, U., Zambounis, J. S., Hao, Z., Iqbal, A. & Dubas, H. European Patent Application No. 0 654 711 A1 (1996).
- Hao, Z., Zambounis, J. S. & Iqbal, A. US Patent No. 5,561,232 (1996).
- Lehn, J.-M. *Angew. Chem. Int. Edn Engl.* **29**, 1304–1319 (1990).

Photosynthesis or planktonic respiration?

The nutrient-poor (oligotrophic) regions of the open ocean cover 30% of the Earth's surface. Microscopic plants (phytoplankton) living in this habitat account for about 10% of global CO₂ fixation¹. Most of this organic production is rapidly respired within a microbial food web dominated by photosynthetic bacteria, heterotrophic bacteria and small (<5 µm diameter) protozoa and algae, with a small amount (about 1% of global CO₂ fixation) being exported from nutrient-poor upper ocean regions to the interior of the ocean^{2,3}. Thus, on average, photosynthesis exceeds respiration in the sunlit 'euphotic' zone of the oligotrophic ocean. This view appears to have been challenged by del Giorgio and co-workers⁴.

Although their conclusion that "...bacterial respiration is generally high, and tends to exceed phytoplankton net production in unproductive systems (less than 70 to 120 µg carbon per litre per day)" is likely to be valid for lakes, it is unlikely to hold in the oligotrophic ocean. An excess of net photosynthesis over respiration is evident in seasonal accumulation of O₂ in near-surface waters⁵ at those times when nutrient supply from the interior of the ocean is cut off by vertical stratification. Net photosynthesis is also evident in downward transport (export) of particulate² and dissolved organic matter³ from the euphotic zone to the interior of the ocean.

The conclusion of del Giorgio *et al.* that the oligotrophic ocean euphotic zone is net heterotrophic was based on analysis of estimates of bacterial respiration and primary production taken from the literature. But a systematic bias in methodology, possibly the consideration of phytoplankton respiration, may account for such a discrepancy.

Both bacterial biomass and bacterial respiration in oligotrophic ocean environments was probably overestimated by del Giorgio *et al.* Photosynthetic picoplankton can be mistaken for heterotrophic bacteria in conventional epifluorescence counts,

leading to an overestimate of the abundance of heterotrophic bacteria in the most oligotrophic waters. For example, photosynthetic bacteria within the genus *Prochlorococcus* accounted for 31% of total bacterial counts (upper 200 m) in the oligotrophic North Pacific, and the biomass of photosynthetic picoplankton exceeds that of heterotrophic bacteria⁶.

del Giorgio *et al.* relied on dark O₂ consumption in the <0.8–2-µm size fraction to estimate bacterial respiration. But in fact phytoplankton are likely to contribute significantly to dark oxygen consumption in this size fraction. Respiration is strongly correlated with growth rate in phytoplankton. For phytoplankton with a specific growth rate of 0.7 d⁻¹ typical of open ocean⁷, a specific respiration of 0.4 d⁻¹ is anticipated⁸. This estimate of phytoplankton respiration can be compared with a calculation of the specific bacterial respiration rate of 0.4 d⁻¹ based on a growth rate⁹ of 0.1 d⁻¹ and growth efficiency⁴ of 0.2. Given equal biomass of heterotrophic bacteria and picophytoplankton⁶, it is likely that picophytoplankton account for about 50% of oxygen consumption in the <0.8–2-µm size fraction of oligotrophic ocean waters. As a consequence, heterotrophic bacterial respiration may have been overestimated by 200% in such waters. del Giorgio *et al.* compared their estimate of "bacterial" respiration with net photosynthesis. As the contribution of phytoplankton to respiration is already included in the net photosynthesis measurement¹⁰, del Giorgio *et al.* could have counted the contribution of picophytoplankton to respiration twice in their analysis.

In their paper, del Giorgio *et al.* assumed a respiratory quotient (RQ) of 1.0 CO₂ evolved per O₂ consumed to convert bacterial respiration from oxygen to carbon equivalents. This RQ is unconstrained by observations for open ocean bacteria, but it is likely to be an overestimate of CO₂ evolution from respiration of organic matter that includes proteins and lipids in addition to carbohydrates. Oceanographers typically use a photosynthetic quotient of 1.25 O₂ evolved per CO₂ assimilated to convert measurements of primary production from carbon to oxygen equivalents¹¹ for comparison with respiration rates measured as O₂ consumption. Multiplying the 200% overestimate of heterotrophic bacterial respiration by a photosynthetic quotient of 1.25 leads to an error of 250%.

Cross-system analyses, such as that undertaken by del Giorgio and co-workers, are necessary for clarifying our understanding of the relationship between heterotrophic and photosynthetic metabolism in aquatic systems. However, the conclusion of net heterotrophy of unproductive oceanic systems needs to be tempered, given uncer-

tainty in phytoplankton respiration rates or other possible systematic errors in bacterial respiration and primary productivity measurements.

Richard J. Geider

Marine Biological Association of the United Kingdom,

The Laboratory, Citadel Hill,

Plymouth PL1 2PB, UK

e-mail: rdg@wpo.nerc.ac.uk

- Behrendfeld, M. J. & Falkowski, P. G. *Limnol. Oceanogr.* **42**, 1–20 (1997).
- Michaels, A. F., Bates, N. R., Buesseler, K. O., Carlson, C. A. & Knap, A. H. *Nature* **372**, 537–540 (1994).
- Carlson, C. A., Ducklow, H. W. & Michaels, M. F. *Nature* **371**, 405–408 (1994).
- del Giorgio, P. A., Cole, J. J. & Cimleris, A. *Nature* **385**, 148–151 (1997).
- Spitzer, W. S. & Jenkins, W. J. *J. Mar. Res.* **47**, 169–196 (1989).
- Campbell, L., Nolla, H. A. & Vulot, D. *Limnol. Oceanogr.* **39**, 954–961 (1994).
- Laws, E. A., Dittulio, G. R., Carder, K. L., Betzer, P. R. & Hawes, S. *Deep-Sea Res.* **37**, 715–730 (1990).
- Geider, R. J. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 333–360 (Plenum, New York, 1992).
- Fuhrman, J. A., Sleeter, T. D., Carlson, C. A. & Proctor, L. M. *Mar. Ecol. Prog. Ser.* **57**, 207–217 (1989).
- Williams, P. J. LeB. *et al. J. Plankton Res.* **18**, 1961–1974 (1986).
- Grande, K. D. *et al. Deep-Sea Res.* **36**, 1621–1634 (1989).

del Giorgio and Cole reply — Geider raises two main issues: first, he points to substantial evidence suggesting net autotrophy in surface ocean waters; and second, he argues that the estimates of bacterial respiration we used in our paper¹, which are based largely on measurements of oxygen consumption in size-fractionated samples, could overestimate the contribution of bacteria.

The surface waters of the ocean, taken as a whole, are probably net autotrophic, for the reasons Geider states as well as others. We were careful to state our conclusions about open ocean systems to reflect this view. In fact, we suggested several mechanisms, including temporal lags and export of detrital organic carbon from more productive marine areas, that would be consistent with both our analysis and the idea of oceanic surface waters being net autotrophic. Thus, we do not disagree with Geider about the net autotrophy in oceanic surface waters.

The second issue raised by Geider is whether the estimates of respiration are representative of the metabolic rates of heterotrophic bacteria in the open seas. Geider argues that they are not, because of the possible contribution of minute autotrophs to the small-size planktonic fraction. From the point of view of the organic carbon balance, it does not matter if some of the respiration is due to the contribution of autotrophs. If one accepts the rate measurements, the destruction of organic matter by respiration is still larger than the production by photosynthesis. Geider estimates autotrophic and heterotrophic picoplankton respiration based on typical growth and specific respiration rates of these organisms in the sea. As part of this calculation, he

uses the 20% bacterial growth efficiency that we report¹. We derived this efficiency precisely from the respiration measurements which Geider criticizes in his comment.

If one accepts this 20% growth efficiency, as many do^{2–5}, then bacterial respiration is four times larger than bacterial production. Let us combine, then, this value of bacterial growth efficiency with published data on bacterial production and primary production for the world's oceans to derive an independent figure for bacterial respiration. The most comprehensive review is that of Ducklow and Carlson⁶. Using these data, the resulting bacterial respiration represents on average 102 and 110% of net primary production in open ocean and coastal marine systems, respectively. Stated more broadly, if bacterial production is ~30% of net primary production, as has been widely reported⁷, when bacterial growth efficiency is 23% or lower, bacterial respiration will exceed primary production. Possibly net heterotrophy is more widespread in the oceans than is currently accepted, or bacterial growth efficiency is much higher than 20%, or most measurements of bacterial production are serious overestimates, relative to primary production.

Paul A. del Giorgio

Jonathan J. Cole

Institute of Ecosystem Studies,
Millbrook, New York 12545, USA

1. del Giorgio, P. A., Cole, J. J. & Cimbleris, A. *Nature* **385**, 148–151 (1997).
2. Kirchman, D. L. *Nature* **385**, 121–122 (1997).
3. Jahnke, R. A. & Craven, D. B. *Limnol. Oceanogr.* **40**, 436–441 (1995).
4. Kirchman, D. L., Suzuki, Y., Garside, C. & Ducklow, H. W. *Nature* **352**, 612–614 (1991).
5. Kristensen, K., Nielsen, H., Reimann, B. & Furrman, J. A. *Microb. Ecol.* **24**, 145–160 (1992).
6. Ducklow, H. W. & Carlson, C. A. *Adv. Microb. Ecol.* **12**, 113–181 (1992).
7. Cole, J. J., Findlay, S. & Pace, M. *Mar. Ecol. Prog. Ser.* **43**, 1–10 (1988).

Lifetime of plasma cells in the bone marrow

Immune protection is based on long-lived memory cells and effector cells, which are either cytotoxic or secrete antibodies. The lifespan of these effector cells has not so far been determined. Here we show that antibody-secreting plasma cells from bone marrow are as long-lived as memory B cells.

After immunization with T-cell-dependent antigens, specific antibody titres are often stable for long periods. As the half-life of antibodies is short¹, the observed antibody titre must be maintained by antibody-secreting plasma cells. In the bone marrow, antigen-specific plasma cells can be detected for at least one year after immunization²,

but whether these plasma cells are continuously generated by B cells that are activated by persisting antigens, or whether they are long-lived, is not known³. Analysis of proliferation of total plasma cells in bone marrow over a 10-day period has suggested that most plasma cells may have a lifespan of only a few weeks⁴. In splenic foci, plasma cells that are derived from B cells after primary immunization die by apoptosis after a few days⁵.

We investigated the lifespan of murine bone-marrow plasma cells in a secondary immune response to ovalbumin. We used cellular-affinity matrix technology⁶ to identify, isolate and analyse IgG1 antibody-

secreting cells (Fig. 1a). Such cells secrete ovalbumin-specific IgG1 antibodies when isolated and cultured (about 0.3–0.5 ng per cell per day; data not shown), and could be identified by intracellular binding of ovalbumin and low expression of surface immunoglobulins.

After secondary immunization with ovalbumin, the absolute number of these plasma cells in the bone marrow (about 5×10^{-4} , ~75,000 per mouse), as determined by flow cytometry, and the specific antibody titre (>1 mg per ml serum) remain constant for more than 3 months, whereas in the spleen, the number of plasma cells drops from a peak of 150,000 per

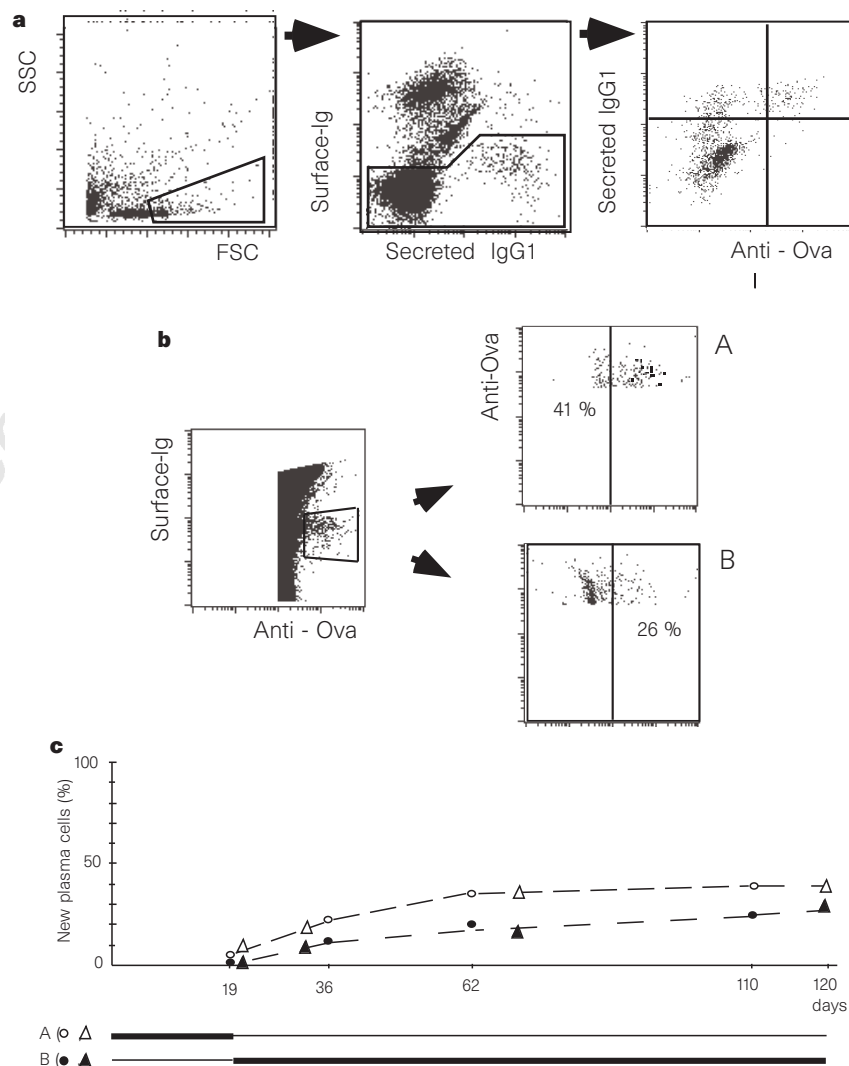


Figure 1 Lifetime analysis of antibody-secreting plasma cells. BALB/c mice were immunized with 0.1 mg alum-precipitated ovalbumin (Ova), and boosted 32 days later with 0.1 mg soluble ovalbumin. Femoral bone marrow of three mice was pooled for each stage of the analysis. **a**, Identification of IgG1-secreting, ovalbumin-specific plasma cells in the bone marrow as blast (forward scatter/side scatter, FSC/SSC), low surface immunoglobulin, high secreted IgG1, intracellular anti-ovalbumin-positive cells, 12 days after boost, with analysis gates as indicated. **b**, Analysis of BrdU incorporation as a measure of DNA synthesis by these bone-marrow plasma cells, 110 days after secondary immunization. BrdU incorporation was determined for group A and B plasma cells (see **c**), by staining with fluorescein-labelled anti-BrdU, with evaluation gates set according to controls (not shown). **c**, Replacement kinetics of ovalbumin-specific bone-marrow plasma cells after ovalbumin boost (day 0). BrdU feeding periods (1 mg per ml drinking water) are indicated by thickened bars. Frequencies of BrdU-negative (A) and BrdU-positive (B) bone-marrow plasma cells. Two experiments are indicated by circles and triangles. (Details of methodology are available from the authors on request).

mouse at day 6 to below 10,000 per mouse from day 11 onwards (data not shown). Therefore, just two weeks after secondary immunization, the majority of specific plasma cells reside in the bone marrow, as expected⁷.

We determined the life-span of ovalbumin-specific plasma cells residing in the bone marrow using bromodeoxyuridine (BrdU) incorporation as a measure for DNA synthesis, a method that had already been used successfully to measure the life-span of memory B cells⁸. We immunized and boosted two groups of mice with ovalbumin. Group A was given BrdU in drinking water for the first 19 days after secondary immunization. For group B, feeding with BrdU started at day 19 after the secondary immunization and was maintained for more than 90 days. We measured BrdU incorporation by ovalbumin-specific plasma cells from each group at various times (Fig. 1b, c). Nearly all of the specific plasma cells developed within two months of secondary immunization. Later, only very few additional plasma cells were found to be labelled in group B, or unlabelled in group A. Group A consistently shows 5–15% more 'new' plasma cells, which probably represent the unlabelled plasma cells from the primary immunization, so they are really very 'old' plasma cells. Apparently, these plasma cells had not proliferated on secondary immunization with ovalbumin. Between primary and secondary immunization, the specific serum titre increased by 10–40 times, corresponding to the 20-fold increase in bone-marrow plasma cells.

To conclude, more than 60% of the plasma cells at day 110 after the booster injection lived for more than 90 days without DNA synthesis. They may live longer, as suggested by the asymptotic slope of the incorporation curve. Our data provide clear evidence that most plasma cells, present in constant numbers in the bone marrow, are long-lived and not derived from the differentiation of proliferating, activated B cells. These plasma cells may be as long-lived as memory B cells. Our result has implications for the design of vaccination strategies and may aid the understanding of allergy and humoral autoimmunity.

**Rudolf A. Manz, Andreas Thiel
Andreas Radbruch**

*Institute for Genetics, University of Cologne,
D50931 Köln, Germany,
and Deutsches Rheumaforschungszentrum Berlin,
Hannoversche Strasse 27, D10115 Berlin, Germany
e-mail: radbruch@drfz.de*

5. Smith, K. G., Hewitson, T. D., Nossal, G. J. & Tarlington, D. M. *Eur. J. Immunol.* **26**, 444–448 (1996).
6. Manz, R. A., Assenmacher, M., Pflüger, E., Miltenyi, S. & Radbruch, A. *Proc. Natl Acad. Sci. USA* **92**, 1921–1995 (1995).
7. Benner, R., Hijmans, W. & Haaijman, J. J. *Clin. Exp. Immunol.* **46**, 1–6 (1981).
8. Schitte, B. & Rajewsky, K. *Nature* **346**, 749–751 (1990).

Importance of ancestral DNA ages

Recent publications consider the age of a common ancestor of samples of human DNA sequences^{1–9}. In particular, variation in Y chromosomes has been interpreted by different authors to give very different estimates of the time to the most recent common ancestor of the sample^{1–8}. The cause of the differences is that the data are being interpreted by various authors in terms of their preferred models of human demography. The data are so lacking in power that these estimates depend hardly at all on the data, and almost entirely on the demographic model assumed. However, if we knew the demography of early humans, we would have no real interest in the time to the most recent common ancestor of an individual gene, as the time is important only because it tells us about the demography.

Generally, the data are a sample of sequences, D . Assumptions about the neutral mutation rate, μ , and a demographic model, N (often including neutrality and a constant effective population size) are used to calculate $P(T|D, N, \mu)$, the probability that the sequences had a time to most recent common ancestry of T , conditional upon D , N and μ . From these, one can calculate an expected value of T , $E(T)$.

Different values of $E(T)$ originate from differing choices of N . A clear example concerns analyses of the data of Dorit *et al.*¹, who found no variation in a sample of 38 Y chromosomes for a 729-base-pair intron from the *ZFY* gene. They estimated $E(T)$ using two models. One incorporated a constant population size, with all values of T equally likely prior to the data. Another model assumed a population expansion again with a constant prior distribution of T . $E(T)$ values were 270,000 and 27,000 years, respectively. Others sought explicitly to condition on the basis of a demographic model. Fu and Li² considered a constant

population size model, and, depending upon the effective population size (from 2,500 to 30,000) obtained values for $E(T)$ ranging from 92,000 years to 703,000 years. Donnelly *et al.*³ assumed a lognormal prior distribution of population size with a mean of 36,000. $E(T)$ values ranged from 254,000 to 460,000 years, depending on μ . Weiss and von Haeseler⁴ imagined populations exponentially increasing, with various growth rates, creating a range of $E(T)$ values from 17,000 to 286,000 years.

The explanation of this 40-fold range of expected times is that, when data sets lack power, $E(T)$ depends more on N than on D . An extreme example is the analysis by Fu and Li⁵ of the data of Knight *et al.*⁹. The sequence diversity of an Alu insertion at the α -globin-2 locus suggested an $E(T)$ of under 100,000 years. In their re-analysis, Fu and Li calculate $E(T)$ values over 450,000 years. They use a neutral model with a constant 10,000 individuals, such that $E(T|N) \approx 800,000$ years, prior to the data, the lack of power of which leaves $E(T|D, N, \mu)$ high.

The demographic models are assumed to be known without error, and are usually not tested for consistency with the data. Yet T is interesting only for the information that it supplies about N , the demography of the human species. It is N , not T , that can be related to the fossil record, for example. We wish to know human population sizes in the past, and to investigate the possibility of demographic expansions during the recent spread from Africa. Values for times to common ancestry calculated on the basis of an assumption that there has been no expansion are merely confusing.

J. F. Y. Brookfield

*Department of Genetics,
University of Nottingham,
Queens Medical Centre,
Nottingham NG7 2UH, UK*

1. Dorit, R. L., Akashi, H. & Gilbert, W. *Science* **268**, 1183–1185 (1995).
2. Fu, Y.-X. & Li, W.-H. *Science* **272**, 1356–1357 (1996).
3. Donnelly, P., Tavaré, S., Balding, D. J. & Griffiths, R. C. *Science* **272**, 1357–1359 (1996).
4. Weiss, G. & von Haeseler, A. *Science* **272**, 1359–1360 (1996).
5. Fu, Y.-X. & Li, W.-H. *Mol. Biol. Evol.* **14**, 195–199 (1997).
6. Dorit, R. L., Akashi, H. & Gilbert, W. *Science* **272**, 1361–1362 (1996).
7. Hammer, M. F. *Nature* **378**, 376–378 (1995).
8. Whitfield, L. S., Sulston, J. E. & Goodfellow, P. N. *Nature* **378**, 379–380 (1995).
9. Knight, A. *et al. Proc. Natl Acad. Sci. USA* **93**, 4360–4364 (1996).

correction

In "Structures of mollusc shell framework proteins", by S. Sudo *et al.* (*Nature* **387**, 563–564; 1997) the deduced amino-acid sequence of pMSI2 (Fig. 1b) contained an error at position 69. The corrected sequence is reproduced below.

b	
MPKPFVTLASLIVLILASASADGYDDYKKGYSVGYGPGISLGGGLGGGGIIISVGGGGGLGGGLGGGLGCG	70
LVGVGGGGLIGGGFGPRVSGTINAGGGVFASGSLGGLSPAGRGARQAATLSALQIASGRPRVSGVS	140
VGTGGGRAVSGSATPVGGFGVPYGGYGYNYGVPSYGVGLPSYGVSLPSYGVGLGGYGGYGLDLASF	210
QGSTYGNLATGQINTAVVAFHMAVILLESMEAESDTEVDTEMDSEEDMESEEDTESEEDTESEEDTESE	280
DMESEEDMDSESSVVDQVSMVYPNHFTEGDVLFQVRLQELEFPALALVSVVLE	334

1. Vieira, P. & Rajewsky, K. *Eur. J. Immunol.* **18**, 313–316 (1988).
2. Slika, M. K., Matloubian, M. & Ahmed, R. J. *Virol.* **69**, 1895–1902 (1995).
3. MacLennan, I. C. M., Liu, Y. J. & Johnson, G. D. *Immunol. Rev.* **126**, 143–161 (1992).
4. Ho, F., Lortan, J. E., MacLennan, I. C. M. & Khan, M. *Eur. J. Immunol.* **16**, 1297–1301 (1986).

Inventor who turned his back on riches

Wizard: The Life and Times of Nikola Tesla – Biography of a Genius

by Marc J. Seifer

Birch Lane: 1996. Pp. 542. Distributed in the United Kingdom by Biblios POS Ltd at £27

L. Pearce Williams

Americans have long lionized those whose lives and achievements have not followed the general trajectory of success in the United States. Thomas Alva Edison and Nikola Tesla represent two sides of this phenomenon.

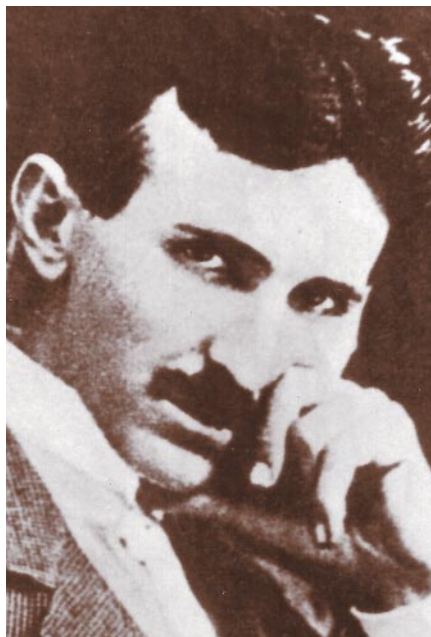
Edison was the poor boy with only a rudimentary formal education who dazzled his contemporaries by his inventions and his fortune. Tesla was the exotic immigrant who turned his back on millions that could have come his way if he had capitalized on such inventions as the polyphase induction motor which could run smoothly off alternating current. It was this current that later revolutionized the distribution of electrical power throughout the world.

Both men combined fertility of imagination with extraordinary insight into electrical processes and the ability to create practical examples from their inventions. And both were dubbed “wizards” because they seemed to the layman to be able to pull new physical effects and inventions apparently and literally out of thin air, or perhaps I should say the ether.

All this was appealing to the American public, both during the men's lifetimes and after their deaths, for it represented a basic American myth. Success could be achieved almost uniquely in America where class, wealth and formal education could be disregarded in the race for fame and riches.

Many books about both men have been written to inspire and instruct the young and to swell patriotic pride in adults. Such books were aimed at a popular audience and could not linger over the technical dimensions of the men's achievements. And beyond the technical details were the industrial and financial aspects of the inventions. Marc Seifer has attempted to put the whole story together in the case of Tesla.

Seifer's autobiographical notes do not inspire confidence in his ability to deal with such a complex scientist and engineer as Tesla. Seifer holds a doctorate from the Saybrook Institute in San Francisco, although his specialist area is not given. His dissertation was entitled “Nikola Tesla: Psychohistory of a Forgotten Inventor”. He has published articles and produced a video on aspects of Tesla's life, many of which locate Tesla on the borders of what most scientists consider to be pseudoscience. Seifer's specific area of competence appears to be graphology, a “science” that he uses as an aid



Tesla: misunderstood genius?

to his interest in psychology, which he teaches. There is no indication that he has any training in the history of science or the history of technology. His book clearly reflects his background.

The introductory address to the volume that commemorated Tesla's eightieth birthday in 1936 cites Tesla's three major scientific and technological contributions. The first was his invention of the first polyphase generator, which permitted the transmission of electrical power over long distances and, at the other end of the wire, the invention of the induction motor, which converted electricity into useful work in a thousand applications.

The second achievement was Tesla's work with high-frequency currents and associated phenomena. The application singled out as most important then was in medicine under the name of “diathermy” or, as it was known in France and Europe, “d'arsonvalisation”. The third contribution cited was his ideas on radio transmission which remained without full realization at the time of his death.

Beyond these inventions were Tesla's manifold interactions with major US business and technological interests — Edison and Westinghouse are only the most famous of the entrepreneurs with whom he dealt and often battled. Finally, Tesla was a popular figure with the US press because of his willingness to make wild and sometimes not-so-wild predictions about what the future would bring.

It was Seifer's obvious intention to bring all these aspects of Tesla together. He has done a very impressive job of collecting a massive amount of documentation of original and

secondary sources, all of which are in English with two possible exceptions. I could not find citations to any of the French or German articles that appeared in celebratory volumes. I cannot help wondering if something has been missed. Still, this is the most complete list of source materials on Tesla. The letters in the J. P. Morgan collection, in particular, shed considerable new light on Tesla's connection with Morgan and his enterprises.

Armed with these sources, Seifer sets out to give as complete an account as he can of Tesla's life, personality, psychological strengths and weaknesses, ambitions, emotions, social relations, persecutions, misunderstandings, victimizations and anything else that impinged upon his hero. Nothing seems to have been put aside as irrelevant or repetitious or simply unbelievable.

The reader is engulfed in a mass of detail, much of which seems merely confusing. Why, for example, is Tesla almost always described as a mathematician when no evidence is presented that his mathematical competence was at the level of contemporary mathematicians or even his fellow “electricians”? Oliver Heaviside comes immediately to mind. Why do we need to know that H. P. Brown, whom Edison invited to his laboratory to experiment on killing animals with alternating current, “lived on Fifty-fourth Street”? There are hundreds of such asides that serve only to divert the author from his main account.

The work is liberally sprinkled with quotations, a practice acceptable in biographies because they preserve both the meat and the flavour of contemporary discourse. But, when these passages are constantly interrupted by ellipses, the chain of thought is broken and the reader ends up seeing spots before the eyes. The practice is not only annoying but also suspect, because with unlimited ellipses a passage can be made to say almost anything the author wishes, including the opposite of what it really does say. Nor does Seifer's style impress. Here is a sample: “A sensualist, White became one of the leading choreographers of the ambience of the percolating metropolis.”

All this is combined with the author's conviction that Tesla was always a victim of malicious and jealous associates, that he was always right, and that others stole his ideas shamelessly, even when he offered them freely. In short, Seifer thinks Tesla was a misunderstood genius exploited by his friends, enemies and anyone else who crossed his path. Paranoia can be fun, but it is bad history.

I have questions about the reliability of the hundreds of footnotes. It was impossible to check them all, so I chose one at random. One paragraph begins with the statement that

"Tesla unveiled the first radio tube in this second month of 1892". This is followed by a description of electrical discharges in a vacuum and a quotation with a footnote. But the source cited does not refer to the material in this paragraph, nor is the quotation there.

I wish I could have written a positive review. In this 542-page work there is a good 200-page book struggling to get out. All the requisite material is there. What is needed is a ruthless editor and the two professional and critical readers that most academic publishers require. □

L. Pearce Williams is at 207 Iradell Road, Ithaca, New York 14850, USA.

House of cards

The Fabric of Reality

by David Deutsch

Allen Lane: 1997. Pp. 390. £25. To be published in the United States on 1 August at \$29.95

Peter T. Landsberg

The great outpouring of 'popular science' books continues unabated. Often they deal with modern physics and provide an introduction to recent progress. Sometimes they suggest a new philosophy or world view. This book is in the second category. Knowledgeable, fiercely outspoken and quite partisan, David Deutsch aims to introduce us to his view of reality. He has written about this

before, but this book presents a more systematic development.

His four strands are quantum theory, the theory of computation, the theory of evolution and the theory of knowledge. He concludes that scientists have not taken their own theories sufficiently seriously; they are instead "clinging irrationally to what could be called 'paradigms'". For example, optical interference experiments involve photons we can detect, and others whose presence we infer merely from their interference effects on the observable photons ("tangible" and "shadow" photons respectively). Take this seriously, Deutsch advises, and call the tangible particles our universe and each shadow particle a part of a parallel universe.

This raises a terminological point. Deutsch drops the term "universe" as a description of the whole of physical reality, because what we see and study would then be merely a portion of it, and so would require a new name. He uses "universe" to denote only the part of physical reality that we normally study. And the rest — the physical reality as a whole? Well, just call that a "multiverse", and we have got our nomenclature. There seems to be no difference between the multiverse and the notion of parallel universes.

Things analogous to each other are supposed to happen in the parallel universes. In one of his few lighter touches, Deutsch speaks of variants of himself in these parallel universes: "Many of those Davids are at the moment writing these very words. Some are

putting it better. Others have gone for a cup of tea."

How can one understand these ideas? When an electron goes through two slits (in the double-slit experiment) there results a superposition of states, and one might talk about a plurality of potential electron states. The next step is to associate possible worlds in which these states are realized individually. This is done here. But it will come as no surprise that there is resistance to the phrase "many worlds", for it can conjure up "ontological profligacy gone mad".

Another terminological point, not made by the author, is elementary but important. When physicists talk about "the universe" that occurs in their theories, they always mean some model universe, not the actual universe we observe, to which the model is, at best, a good approximation.

A reader may well want to hurry on to quantum computers, which are associated with some seminal papers by Deutsch. Indeed, here we find a reference to Peter Shor's algorithm — a way of factorizing numbers of the order of 1 followed by 500 zeros. These are huge numbers, if we recall that the number of particles in the observable universe is estimated to be 'only' 1 followed by 80 zeros. The factorization, asserts Deutsch, uses the resources of the parallel universes. His challenge is: "Explain the algorithm on a single-universe world view." I hope that computer scientists can and will take up this challenge.

Here, as in other places, I was disappointed that the author could not give a briefer and more incisive introduction to basic ideas. Must one really still also go to articles in *Science* or *Physics World* to find out about quantum computers when one has this book? I have a feeling that Deutsch is so keen to propagate his ideas that he sometimes neglects a more straightforward exposition. He does not warn us, for example, that there could be obstacles to the construction of quantum computers that may not be overcome for many years. The main difficulties will arise from the nature of the error correction needed and from quantum mechanical decoherence problems.

There is some fun to be had in the chapter on time travel. The most interesting type is travel into the past, and there is still some doubt whether this is 'allowed' by our scientific laws. If it is, one may expect messages from the future. But visitors from the future "cannot know our future any more than we can, for they did not come from there". Deutsch engages us in various speculations, some of which, as he admits, would not be out of place in science fiction.

To whet appetites, here are some memorable quotations: "While most mathematicians and computer scientists take the certainty of mathematical intuition for granted, they do not take seriously the problem of



...nor any drop to drink

Floods come quickly upon villagers in the Ganges and Brahmaputra delta, particularly when dams on the other side of Bangladesh's land borders are opened without warning. People seek refuge on roofs, only to be at risk from deadly bites of

water-shy snakes. From *Delta: The Perils, Profits and Politics of Water in South and Southeast Asia*, a photographic survey by Daniel Schwartz of life in these impoverished yet resourceful delta regions. Thames and Hudson, £28.

In retrospect chosen by Euan Nisbet



Emma
by Jane Austen
(1816)
**The Climate
of London**
by Luke Howard
(1833)

**Austen: acute
observer of
the weather**

Why choose *Emma*? *Emma* is weather. Meteorology shapes the novel. This is a work of science, seamlessly woven into art. *Emma* is perfect. The slightest change would disrupt the book. How did Jane Austen create a work without flaw? Literati will have their explanations and doubtless many learned theses. But, to this illiterate from the numerati, the science perfects the art.

Emma herself is in full bloom, lovely, ready for marriage, as the spring of life passes to high summer. Day by day, the plot twists with the weather report. Is it bright? All is cheerful. Is it drizzling? Misery abounds. Or, beware, is it hot and sultry? Romance and danger loom. No doubt there is a learned tome on this somewhere, but it is a fascinating game to read *Emma* alongside one of the founding texts of meteorology, Luke Howard's *The Climate of London*.

Emma is set not far from London, near Cobham in Surrey, perhaps at Painshill, where the eccentric Mr Hamilton, related by marriage to Admiral Nelson's Emma, had created an experimental garden-farm, a ruined 'abbey' and artistic 'mill'. There was "a sweet view, sweet to the eye and the mind". The gardens are now modernized by electricity pylons so obtrusive that the observer may be permitted to suspect the hand of a planner.

Howard, a chemist and close friend of John Dalton, named the clouds: stratus, cumulus, cirrus, nimbus. He helped to lead the Bible

Society and the fight against slavery. After the devastation of the Napoleonic wars, Howard and his friends in the United Kingdom and United States collected a vast sum, equal to many millions today — and took it to help the distressed people of Germany. Even Goethe, who wrote *In Honour of Howard*, addressed him as 'master'.

On the warm evening of 22 July 1813, Howard records his visit to Alton, Hampshire. As he travelled through Chawton, just before Alton, he would have passed before Austen's dining room window, the outlook of one who was his equal in meticulous observation. Whether they met that day we do not know but it seems possible. Howard was a campaigning celebrity with links to the Lloyd and Barclay families, Quaker bankers. There were Barclays in Alton, and Austen's brother was a banker. After this time, Austen's letters seem full of weather.

Austen wrote *Emma* in 1814–15. It is nice to imagine that the crux of the book, the trip to Box Hill, dates from summer 1814. The lesser details may have been filled in as she wrote. Suppose then that the book records the weather of summer 1814 and winter 1814–15, day by day as she wrote, although the calendar may be 1813–14, when she began the plotting. With these assumptions, the course of the book fits beautifully with the weather recorded in *The Climate of London*. If so, the story may begin on 25 September, pass through autumn to snow at Christmas (now a rare event, but it did occur at Christmas 1814), then to a post-Christmas period between frost and thaw (32–41 °F in Howard's record), and the late winter weather of early 1815.

The crisis in the book occurs just before midsummer's day. Austen makes the fascinating observation of an "orchard in blossom", her famous 'error'. What are apple trees doing in flower in mid-June? But is this error — or clue? The weather was unusual in 1814. The annual mean temperature was one of the coldest in Howard's record, and in May and June the means were colder than 1816, the 'year without a summer' after the eruption of the Tambora

volcano in what is now Indonesia. In the cool spring of 1996, mild in comparison to 1814, apple trees flowered as late as early June.

Perhaps Austen herself saw apple blossom on two hot days, 14 June (85 °F) and 15 June (78 °F), at Painshill and Box Hill. Then the weather broke. On 21–23 June, Howard notes that a "fire in the grate has been again acceptable", an observation worthy of Mr Woodhouse himself. Only as June ended did summer reappear. In July came clouds of uncommon beauty. In *Emma* "it cleared; the wind changed into a softer quarter; the clouds were carried off, the sun appeared; it was summer again". Did Mr Knightley come to call on Wednesday 6 July 1814?

Is it presumptuous to attempt to match the weather to the novel? Possibly — an author has the right of imagination. But Austen is accurate. If she says the orchard was in bloom, then it surely was in bloom. Her meteorological sense is acute, accurately recording the passage of fronts. The perfection of the book comes from the quality of the observation; the science makes the art. Each graduate student should be set to read an Austen novel before starting a thesis.

Austen painted in miniature on two inches of ivory, but she wrote on large issues. In 1772, Judge Mansfield's far-reaching judgment ended slavery in England, helping to create modern individual liberty; in *Mansfield Park* the rich, cultured society can save itself only if it is cleansed of its Caribbean evil. The Prince Regent, to his credit, greatly liked that book. He would have seen the point. We should not forget it today. *Sanditon*, her last work, has a black subsidiary heroine: was Austen going to tackle racism?

In her tiny theatre, Austen took on the greatest of themes. Is it too much to see in her masterpiece, *Emma*, an allegory on *Nature* herself? Perhaps this is imagining too far — or may it be allowable in this journal? But it is a good excuse to follow her commendation and pour a glass of Constantia in her honour this July.

Euan Nisbet is in the Department of Geology, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, UK.

reconciling this with a scientific world view"; "mathematical knowledge is inherently derivative, depending entirely on our knowledge of physics"; "Plato was a very competent philosopher who believed in telling ennobling lies to the public"; "Let us start by imagining some parallel universes stacked like a pack of cards, with the pack as a whole representing the multiverse".

Even wider horizons are explored in the concluding chapter. Deutsch suggests that his four strands together provide an explanation of reality. The fourth strand is knowledge, "which seems a parochial concept until we consider it from a multiverse perspective". Regarding ourselves as pri-

marily knowledge-creating beings, we would welcome an unlimited supply of energy and hence the availability of an enormous number of computational steps. How can this be achieved?

Here Deutsch makes contact with Frank Tipler's "omega-point" theory. The omega-point is supposed to be the end-point of the gravitational collapses after many cycles. The many-cycle concept does not seem as far-fetched as I once feared. Just before the end-point is reached, the energy generated may well be such that "an infinite number of memory accesses" and an unlimited memory capacity are feasible. So we can then march forward into what future is left

for us. It is a huge extrapolation! I am worried about the impression it creates among sensible nonscientists. There is more to life than the creation of knowledge, as also pointed out by Deutsch (somewhat belatedly) in his exposition.

This is a highly stimulating book, full of ideas and spurring us on to greater efforts of imagination; indeed, it lifts the veil that separates science and science fiction. Some of the arguments are hard to follow and some seem quite unconvincing. But they are always striking. □

Peter T. Landsberg is in the Department of Mathematics, University of Southampton, Southampton SO17 1NP, UK.

Up the cosmos

Hunting Down the Universe: The Missing Mass, Primordial Black Holes and Other Dark Matters

by Michael Hawkins

Little, Brown: 1997. Pp. 278. £18.99. To be published in the United States by Addison-Wesley on 1 October at \$25

William H. Press

Occasionally a book can be judged by its cover. This book's cover is black-on-black and barely decodable if you tip it at an angle to the light. You just know that this is going to be an outrageous book, and the author, the Edinburgh astronomer Michael Hawkins, does not disappoint.

He begins with a fierce attack on the likes of Steven Weinberg, Karl Popper and anyone else with the temerity to suggest that science is an objective search for an underlying reality. Addressing the suspicions of the most vehement critics of organized science, he states that, yes, "science is a social process beset by vested interests and power struggles, and scientific 'truth' is therefore determined by consensus, or even complicity". Strong stuff, this. Wittgenstein and Fred Hoyle also figure prominently, but one has a hard time keeping track of what side each is on. The general idea is that good theories — terrific theories even — get rejected by the "establishment" as they break too many rice bowls.

It is not too much of a surprise to learn that the author has a little theory of his own which, if true, is nothing less than "the most important cosmological breakthrough of the century". Is he pulling our leg, or is there something here to be taken seriously? Or a little of each?

Hawkins believes that small, primordial, omnipresent black holes, each about the size of a basketball and with a mass comparable to that of the planet Jupiter, comprise the cosmological "missing" mass. He believes there are enough such black holes to close the Universe and bring it into consistency with Guth's elegant inflationary model.

Virtually all astrophysicists view this hypothesis as highly improbable, for a variety of solid reasons. But it is not flatly contradicted by existing data. Moreover, the hypothesis does make predictions — quite remarkable predictions — that are subject to verification or disproof. As the hypothetical objects drift randomly across the line of sight between us and any distant quasar, their gravitational fields will momentarily focus the quasar's light in our direction. The result is that quasars should fluctuate in brightness.

Quasars do in fact fluctuate in brightness! All astrophysicists except Hawkins believe that these fluctuations are intrinsic to the quasar and are due to flarings of the unstable disc of gas that makes the quasar shine. Hawkins is

pretty much alone in believing that the fluctuations are the proof of his theory, and that the "establishment", embodied in the Astronomer Royal Sir Martin Rees (an unlikely villain if ever there was one), is prejudicially blind to the truth. So this book is his appeal to that greater court of educated lay opinion.

It's a good yarn. But I think the actual situation is more prosaic and less polarized. Hawkins's papers do pass peer review and are published, in *Nature* no less. His ideas are improbable, but observationally testable. In due course, as more data on quasar fluctuations accumulate, his hypothesis is very likely to be disproved to the satisfaction of any sapient individual. If, on the other hand, his theory is observationally confirmed, then he will without doubt be acclaimed by the establishment, and I will eat this review. (It's brief, just in case.) □

William H. Press is at Harvard College Observatory, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

Men behaving badly

Behavioral Genetics, third edition

by Robert Plomin, John C. De Fries, Gerald E. McClearn and Michael Rutter

W. H. Freeman: 1997. Pp. 367. \$64.95, £26.95

Steven Rose

A basic textbook should act as an accessible, comprehensive and balanced introduction to the field of study, tailored to the needs of not only students but also their teachers, who are usually unlikely to be as well-versed in the topics as the textbook's authors. In innovative or controversial areas, this imposes a severe responsibility on the authors. They may and indeed are likely to have a preferred take on the topics they cover. Nothing wrong with that: treated properly, it provides a refreshingly coherent pathway through often messy data.

But textbook authors who 'take a position' also have a responsibility to the students for whom their book will act as source material: they must make it clear where they are coming from, and at least acknowledge by reference the alternative positions they reject. And I know from experience that this is not always easy.

In few areas of modern science can this stricture be sharper than in the biological — and especially the assumed genetic — causes of human behaviour. This is a field that has been vitriolic with methodological, interpretative and political disputes for its entire century-long history. The authors of *Behavioral Genetics* are no novices in this contentious area. Between them the first three have made substantial contributions to arguments that genetics is important in our understanding of individual differences in human behaviour. The fourth, a social psy-

chiatrist, has over the past decades become increasingly interested in the role of genetics.

Their textbook aims to introduce the methods of quantitative and statistical population genetics, some of which the authors and their colleagues have pioneered, to students of psychology and psychiatry. They offer a quick whip through Mendel and single-gene disorders and 'exceptions' to Mendelism, and a beginner's guide to DNA, before embarking on their main task, staking the claim that quantitative genetics can explain the causes of important features of the human condition.

Unfortunately, the result is less a work of balanced scholarship than of passionate advocacy. Its aim, according to the preface, is to make the case that genetics is central to psychology. In attempting to achieve this goal, it is scrupulous with neither the data or their presentation. Few other textbooks are, for instance, illustrated by pictures of leading advocates in the field, supplemented by 'boxes' of uncritical hagiography, while ignoring — with the exception of the odd disparaging remark — equally distinguished psychiatrists, neuroscientists or geneticists who offer a less genetically determinist view of human behaviour.

Few competent textbooks of psychiatry describe such diagnostic categories as schizophrenia, cognitive, attention-deficit or personality disorders without allowing a shred of doubt as to the validity and stability of the categories offered. Some may, but none should, describe adoption studies as straightforward "experiments of nature" without any social reflection on the non-neutral significance of adoption for both the adoptee and his or her parents.

Indeed, the lack of appreciation of the social context, as if the vagaries of human behaviour were as simple to categorize as those of *Drosophila* in a choice maze, is one of the most depressing — and alarming, considering its intended audience of psychiatry students — features of this book. And few textbooks of genetics offer such adulatory accounts of the validity of quantitative trait locus analyses without at least reflecting on the problems of distinguishing biological reality from statistical artefact.

A textbook should encourage students to think for themselves, to develop a healthy respect for the science introduced and to cultivate a critical approach to assessing the theoretical and experimental validity of its claims. Otherwise students will be merely drilled into the latest fashionable theoretical mould.

In their zeal for pressing their own views, influential though they may be, the authors of this textbook have badly failed their basic educational responsibility. □

Steven Rose is in the Brain and Behaviour Research Group, Open University, Milton Keynes MK7 6AA, UK.

Proximal–distal axis formation in the *Drosophila* leg

Thomas Lecuit & Stephen M. Cohen

European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Limb development requires the formation of a proximal–distal axis perpendicular to the main anterior–posterior and dorsal–ventral body axes. The secreted signalling proteins Decapentaplegic and Wingless act in a concentration-dependent manner to organize the proximal–distal axis. Discrete domains of proximal–distal gene expression are defined by different thresholds of Decapentaplegic and Wingless activities. Subsequent modulation of the relative sizes of these domains by growth of the leg is required to form the mature pattern.

Short-range interactions between cells in adjacent compartments trigger the patterning of the legs and wings of *Drosophila*^{1,2}. These interactions direct expression of the secreted signalling molecules Wingless (Wg) and Decapentaplegic (Dpp) in cells adjacent to the dorsal–ventral (D–V) and anterior–posterior (A–P) compartment boundaries of the wing^{3–9}. The secreted Dpp protein is thought to form a concentration gradient that organizes the A–P axis of the wing by directly specifying distinct domains of gene expression^{10–13}. Similarly, localized expression of Wg in cells adjacent to the D–V boundary is thought to generate a gradient of Wg activity that directly specifies distinct domains along the D–V axis of the wing^{9,14–16}.

In the developing leg, Wg and Dpp have been implicated in both D–V and proximal–distal (P–D) axis formation. In contrast to the wing, Wg and Dpp are both expressed in the leg along the A–P boundary under the control of Hedgehog (Hh)⁴. Wg is expressed ventrally, whereas Dpp is expressed at high levels dorsally and at low levels ventrally (Fig. 1). The asymmetry in Wg and Dpp expression is maintained by mutual repression^{17–19}. Wg is required for specification of ventral cell fates, and Dpp is required for specification of dorsal cell fates^{17,18,20–23}.

The central region of the disc where cells expressing Wg and Dpp

meet corresponds to the most distal tip of the leg. It has been shown that producing an ectopic intersection point between cells expressing Wg and Dpp can induce a supernumerary P–D axis^{17–19,21,22,24}. Here we present evidence that Wg and Dpp act directly to provide positional information along the P–D axis of the leg. The combined activity of both signals is required in a spatially graded manner to define the distinct domains of gene expression along the P–D axis. We also show that the time at which these expression domains are stably defined plays an important role in the patterning of the leg.

Wg- and Dpp-dependent target genes in the leg

The Dachshund (Dac) and Distal-less (Dll) proteins are expressed in distinct domains along the P–D axis of the leg (Fig. 1). *Dll* encodes a homeodomain protein that is required for limb development²⁵. In the mature leg disc, *Dll* is expressed in a central domain that corresponds to the presumptive tarsal segments and distal tibia²². The *dac* gene encodes a nuclear protein required for development of the femur and tibia²⁶. *Dac* is expressed in a ring corresponding to the presumptive femur, tibia and first tarsal segment, but is absent from the more distal tarsal segments of the leg disc (Fig. 1b–e). Although there is little or no overlap between the *Dll* and *Dac* domains at early stages (Fig. 1b), by mid third instar

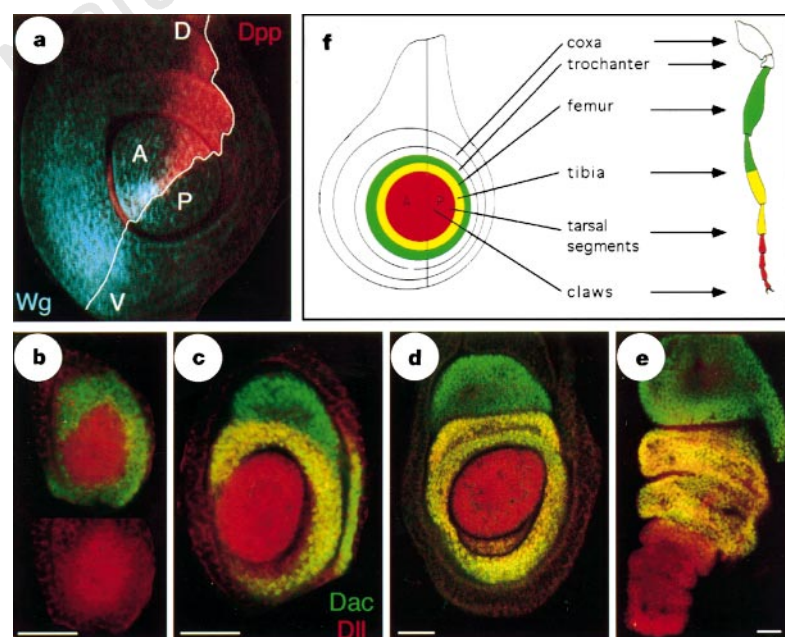


Figure 1 Subdivision of the P–D axis by expression of *Dll* and *Dac*. **a**, Wg protein (blue) and *dpp-lacZ* (red) expression in a third instar leg imaginal disc. A, P, D and V denote anterior, posterior, dorsal and ventral, respectively. The white line indicates the A–P compartment boundary. Wg is expressed by ventral anterior cells; secreted Wg protein can also be seen in the posterior compartment; *dpp-lacZ* expression in dorsal anterior cells is visualized with anti- β -galactosidase; *dpp-lacZ* is also expressed at lower levels ventrally (not visible here). **b**, *Dll* (red) and *Dac* (green) expression in a late second- or early third-instar leg disc (~72 h at 25°C). *Dll* is expressed in a central domain, *Dac* is expressed in a ring. The *Dac* domain overlaps the edge of the *Dll* domain by at most one or two cells. **c**, **d**, Mid and late third-instar discs. Note the overlap of the *Dll* and *Dac* domains (yellow). Scale bars indicate the relative magnification. By mid–late third instar, *Dll* is activated in a secondary domain forming a thin ring at the level of the trochanter (not visible in these optical sections; see Fig. 2d). **e**, Partly everted pupal leg disc showing overlapping *Dll* and *Dac* expression. **f**, Schematic representation of the relationship between the third instar disc and the adult leg. Distal structures derive from the centre of the disc, progressively more proximal structures from a series of concentric rings. The concentric nature of these expression domains is partly obscured by the folding of the disc epithelium.

the combination of Dac and Dll expression defines three regions along the P–D axis (Fig. 1c,d). The relationship between the Dll and Dac expression domains and the segments of the adult leg can be seen in the partly everted pupal leg (Fig. 1e,f).

Dll and Dac are expressed in circular domains centred on the point at which the Wg and Dpp expression domains meet in the leg disc. Two lines of evidence suggest that Dll expression in the centre of the disc depends on the combined activity of *wg* and *dpp*: (1) Dll is not expressed in *wg* or *dpp* mutant discs²²; and (2) misexpressing Wg directs ectopic Dll expression only on the dorsal side of the disc where Dpp is expressed, whereas misexpressing Dpp or ectopically activating the Dpp signal transduction pathway directs ectopic Dll expression only on the ventral side of the disc where Wg is expressed²² (Fig. 2).

Wg and Dpp act directly to induce Dll

We can consider two mechanisms by which Wg and Dpp could induce Dll. Wg and Dpp might act locally in the centre of the disc to induce expression of a third signalling molecule that works at a distance. In this case, cells in which the Wg and Dpp signal transduction pathways are activated would be expected to generate a signal that induces Dll expression non-autonomously in nearby cells. Alternatively, the two signals might both act directly at a distance, in which case Dll expression would be restricted to cells in which both signal transduction pathways are activated. To distinguish between these possibilities, we compared the effects of Dpp-expressing clones with clones in which the Dpp signal transduction

pathway is constitutively activated by expressing a ligand-independent form of the Dpp receptor, Thick-veins (*Tkv**; the asterisk denotes the activated receptor)^{11,12}. Dpp-expressing cells act non-autonomously to induce Dll expression in cells surrounding the clone (Fig. 2a). In contrast, Dll is induced only in cells expressing *Tkv**, but not in neighbouring cells where the pathway has not been activated (Fig. 2b,c). Similarly, although Wg-expressing cells act non-autonomously to induce Dll (Fig. 2d), ectopic expression of Dll and Dac are restricted to *zw3* mutant cells in which the Wg signal transduction pathway is activated cell autonomously (Fig. 2e,f). These results indicate that the long-range effects of Wg and Dpp cannot be due to induction of another secreted signal in cells in which both pathways have been activated. Rather, Wg and Dpp seem to act directly and at long range to induce Dll. Evidence that Dac is regulated similarly is presented below.

Transient requirement for Wg and Dpp signals

Although Wg and Dpp signalling seems to be required to induce Dll and Dac expression, we found that both target genes are expressed normally in clones of cells mutant for a null allele of *dishevelled* (*dsh*), a component of the Wg signal transduction pathway (Fig. 3a,c). This contrasts with the wing, in which *dsh* clones cause a strong reduction in Dll expression^{15,16} (Fig. 3b). The *dsh* clones were induced after the onset of Dll and Dac expression in these experiments, raising the possibility that Dll and Dac do not require continuous input from Wg once expression has been induced in the leg. To test this proposal we examined discs in which Dll is

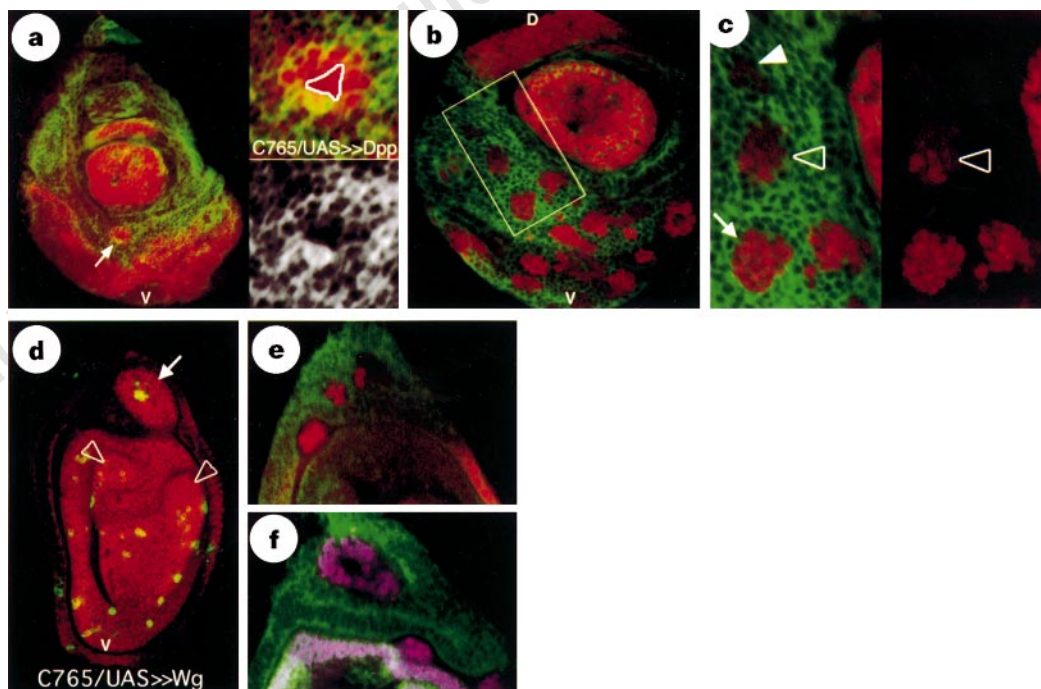


Figure 2 Wg and Dpp do not act by relay of another signal to induce Dll. **a–c**, Comparison of clones expressing Dpp and *Tkv** 24 h after clone induction. **a**, UAS >> Dpp expressing clone (arrow). The clone is marked by the absence of CD2 expression (green). Dll (red) is expressed both inside and outside the clone (outlined in the higher-magnification image, right). **b**, Leg disc with multiple clones expressing UAS >> *Tkv**. V, ventral; D, dorsal. **c**, Detail of the boxed region in **b**. Note that the panel was rotated so that ventral is at the bottom. Arrow, ventrally located clone showing Dll expression exclusively inside the clone; open arrowhead, lateral clone in which the Dll level decreases from ventral to dorsal; filled arrowhead, dorsal clone with little or no Dll expression. **d**, Leg disc with multiple Wg-expressing clones (green, visualized by expression of the flu epitope tag)

examined 36 h after clone induction. Arrow, dorsally located clone showing non-autonomous activation of Dll; open arrowheads, overgrowths induced in the Dll expressing region on the dorsolateral sides of the disc which distort the shape of the disc. Ventral and lateral clones do not induce ectopic Dll. **e, f**, Dorsal regions of leg discs with *zw3* mutant clones induced 24–40 h before staining (marked by the absence of lacZ expression, green). Dll (**e**) and Dac (**f**) are expressed autonomously in mutant cells where the Wg pathway is activated; *zw3* mutant clones induced at this stage do not express Wg (data not shown), although clones induced before the start of third instar can induce Wg in the anterior compartment and act non-autonomously¹⁸.

induced *de novo* outside its endogenous expression domain, and investigated whether Dll could be induced in *dsh* mutant cells. We made use of the fact that Hh directs expression of both Wg and Dpp in anterior leg cells, and that Wg signalling interferes with the ability of Hh to induce Dpp in ventral cells^{17,18}. Consequently, Wg-dependent repression of Dpp cannot occur in ventral–anterior clones of *dsh* mutant cells¹⁸. Thus *dsh* clones serve as a source of Dpp and induce a secondary P–D axis, but are themselves unresponsive to Wg. A large *dsh* mutant clone located outside the endogenous Dll domain that has caused an axis duplication is shown in Fig. 3d. Dll is induced in cells adjacent to the clone (arrow) but not in the *dsh* mutant cells (asterisk). Taken together these results indicate that cells need to receive the Wg signal to induce Dll expression, but do not require continuous signalling to maintain Dll expression after its initial induction.

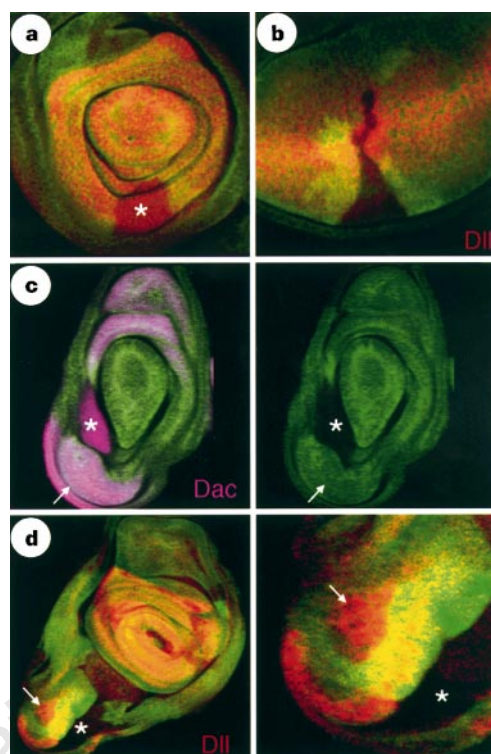


Figure 3 Wg signalling is required for induction, but not maintenance, of Dll and Dac. **a, b**, *dsh*^{VA153} mutant clones induced in second instar larvae (60 ± 12 h; marked by the absence of β-gal expression, green). **a**, Leg disc with a *dsh* clone in the Dll domain (asterisk). Dll expression (red) is normal in the clone. **b**, Dll expression is strongly reduced in a comparable clone in the wing disc. **c, d**, Axis bifurcations associated with *dsh*^{VA153} mutant clones induced in second instar (60 ± 12 h). *dsh* mutant clones in the ventral–anterior region of the disc alleviate Wg-dependent repression of Dpp and induce secondary P–D axes. Axis duplications were identified as morphologically visible outgrowths associated with *dsh* mutant clones on the ventral side of the disc. Clones were marked by the absence of β-gal expression (green). **c**, *dsh* clone (asterisk) located in the Dac expression domain on the ventral side of the disc. Dac expression (purple) is normal in the clone, but is expanded in the induced outgrowth (arrow; compare with wild-type disc in Fig. 1). **d**, *dsh* clone located outside the endogenous Dll expression domain. Dll is not expressed in the *dsh* mutant cells (asterisk) but is induced in adjacent wild-type cells (arrow). A higher magnification of the duplication is shown on the right. The ectopic Dll domain appears red where it overlaps the heterozygous *dsh*^{+/dsh} cells, and yellow where it overlaps the bright green cells carrying two copies of the β-gal marker and of the *dsh*⁺ chromosome. Ectopic expression of Dll or Dac is always associated with obvious axis bifurcation. Comparable results were obtained for Dac expression associated with *dsh* clones in the antenna (not shown).

A similar distinction between induction and maintenance can be made for Dpp signalling. Clones of cells mutant for *Mad* (a component of the Dpp signal transduction pathway^{27,28}) or the Dpp receptor gene *tkv* express normal levels of Dll (Fig. 4a,b) and Dac (Fig. 4c). These results indicate that continuous Dpp signalling is not required to maintain Dll or Dac expression in the leg. We examined axis bifurcations induced by *Mad* mutant clones to look at *de novo* induction of Dll and Dac. *Mad* mutant clones block transduction of the Dpp signal and so allow Hh to induce Wg in dorsal cells (as reported for the Dpp receptors *punt* or *tkv*^{18,19,29}). *Mad* clones induced near the A–P boundary cause axis bifurcation. However, to increase the frequency of axis bifurcation we made clones simultaneously mutant for *Mad* and *pka*. The *pka* mutant cells behave as though the Hh pathway is activated^{30–33}, thereby allowing *Mad pka* mutant clones to express Wg, regardless of

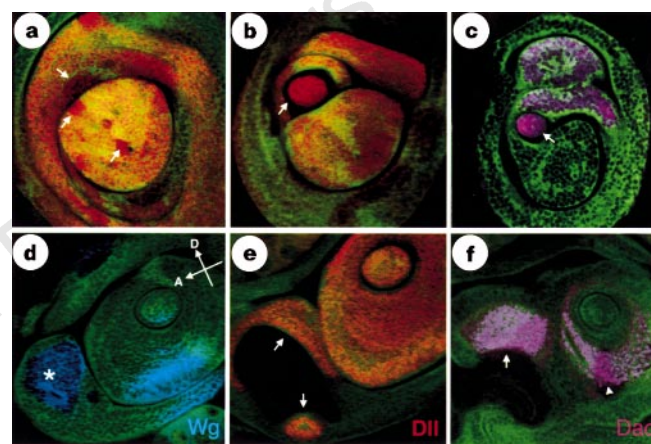


Figure 4 Dpp signalling is required for induction, but not maintenance, of Dll and Dac. **a, b**, *Mad* mutant clones in the leg disc (arrows, marked by the absence of β-gal, green). Dll expression (red) is unaffected in *Mad* clones induced in early third **(a)** or second **(b)** instar larvae (60 ± 12 h). **c**, Dac expression (purple) is unaffected in a *tkv* mutant clone (arrow, induced 60 ± 12 h). **d–f**, *Mad pka* double-mutant clones in the antenna disc (induced 60 ± 12 h; marked by the absence of β-gal, green). Dorsal and anterior are indicated. Clones located outside the Dll and Dac domains are easier to find in the antennal disc, because it is less folded than the leg disc. Identical results were obtained in leg discs, but fewer examples were found. **d**, Wg expression (blue) induced in a *Mad pka* clone far from the A–P boundary (asterisk). **e**, Dll expression (red) is induced in cells adjacent to a comparable clone (arrow), but not in the *Mad pka* mutant cells. **f**, Dac expression (purple) is induced in cells adjacent to a *Mad pka* clone (arrows), but not in the mutant cells. The arrowhead indicates a ventral *Mad pka* clone in the endogenous Dac domain near the A–P boundary. Dac is expressed normally in the clone. No axis duplication results because the clone is located in the region where Wg is endogenously expressed. The Dac domain is narrower than the Dll domain in the antenna, so that Dll expression extends to a more proximal region of the antenna, the reverse of the situation in the leg. Dac is expressed on the side of the clone close to the source of Dpp, whereas Dll is expressed on both sides of the clone.

whether they are within range of the Hh signal (Fig. 4d, asterisk). *Mad pka* mutant clones located outside the endogenous Dll or Dac domains can induce axis bifurcations and expression of Dll and Dac. Dll and Dac are induced only in wild-type cells adjacent to the clone (Fig. 4e,f, arrow), but not in the *Mad pka* mutant cells, which serve as a source of Wg, but are unresponsive to Dpp. These results indicate that, like Wg, Dpp signalling is required for the initial induction of Dll and Dac expression but not for their subsequent maintenance. Thus the requirement for these signals seems to be transient.

Distinct thresholds of Wg and Dpp activity

The spatial domains of Dac and Dll differ along the P–D axis of the leg. Dll is expressed in a distal cap that has its proximal limit in the tibia; Dac is expressed in a ring that has its proximal limit in the femur. We have excluded the possibility that Wg and Dpp induce a new signal that radiates outward from the centre of the disc, and

have shown that both signals are required directly for the induction of Dac and Dll. This raises the possibility that the domains of Dll and Dac expression are defined in response to different threshold levels of Wg and Dpp activity. Evidence has been presented for a graded activity of Wg in the specification of cell fate along the ventral-to-dorsal axis of the leg^{15,17,21}. To assess the effective range of the Wg signal, we examined Dll expression in discs carrying many small *Tkv*^{*}-expressing clones (in which the level of Dpp signalling activity should be uniform). Dll is expressed in *Tkv*^{*} clones located far from the source of Wg (Fig. 2b, double labels with Wg not shown). Dll is induced in all cells of ventrally located *Tkv*^{*} clones (arrow in Fig. 2c). In a ventrolaterally located clone (open arrowhead), Dll is found at a higher level in cells close to the source of Wg than in the cells on the far side of the clone. Clones located more dorsally express little or no Dll. The decrease in Dll expression with increasing distance from the source of Wg suggests that the level of Wg activity is limiting (when there is sufficient Dpp activity). We

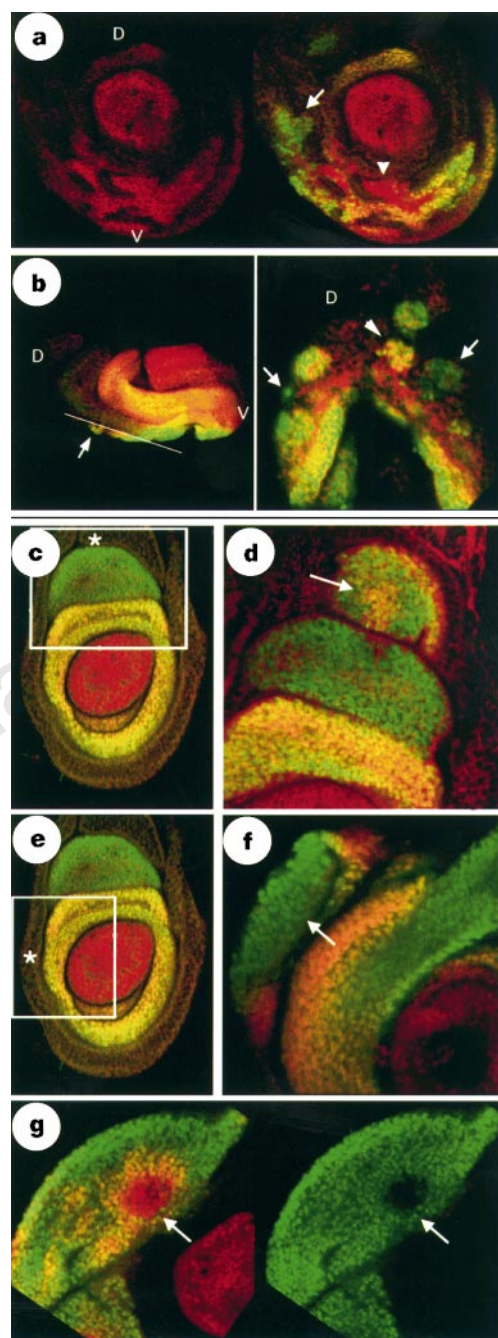


Figure 5 Distinct thresholds for activation of Dll and Dac by Wg and Dpp. **a, b**, Leg discs double-labelled for Dac (green) and Dll (red). **a**, Ventral clones expressing the activated Dpp receptor *Tkv*^{*} express Dll, but not Dac (arrowhead). Dll is shown separately to the left. Dac is activated in more dorsally located clones (arrow). The image is focused on the proximal region of the disc which is overgrown. Only the tarsal region of the normal Dll domain is in the plane of focus; most of the normal Dac domain is not. **b**, Clones expressing an activated form of the Wg pathway component Armadillo. Left, side view of a leg disc showing the approximate plane of the optical section on the disc shown in the right panel. The arrow shows a dorsal clone expressing Dac. Right, Dac and Dll are coexpressed in activated Armadillo clones located near the source of Dpp on the dorsal side of the disc (arrowhead). Dac is expressed alone in clones located more laterally (arrows). The endogenous Dll and Dac domains in the centre of the disc cannot be seen. **c**, Wild-type disc (from Fig. 1d). The asterisk indicates the approximate position of the Wg-expressing clone shown in **d** relative to the endogenous Dll and Dac domains. The box indicates the portion of the disc shown. **d**, Dorsal *Act* >> Wg-expressing clone located outside the endogenous Dll and Dac domains (arrow). Dac (green) is misexpressed in a larger region than Dll (red). The ectopic Dac expression domain may have merged proximally with the endogenous domain (these domains are separated by a deep fold of the disc epithelium). **e**, Wild-type disc. The asterisk indicates the approximate positions of the Dpp-expressing clones in the discs in **f** and **g**. **f, g**, *Tub* >> Dpp-expressing clones outside the endogenous Dll and Dac domains (arrows). **f**, Clone examined 24 h after induction, in which Dac is induced, but not Dll. The expression of Dll near the edge of the Dac domain is the endogenous proximal ring of Dll at the border between femur and trochanter²². **g**, Clone examined 40 h after induction, in which Dll is induced in a subset of Dac-expressing cells and Dac is repressed in the centre of the Dll-expressing region (arrows). Dac is shown separately on the right. The Wg- and Dpp-expressing clones were not directly marked in these experiments because antibodies to Dac and the clone markers are both of mouse origin. Based on the relationship between Dll expression and marked clones in Fig. 2, the Dll expression domain shows the approximate position of the clone, to allow comparison with Dac expression.

propose that a spatial gradient of Wg activity determines both the range over which Dll can be activated and the level of Dll activation. It is possible that the graded Wg activity reflects a gradient of Wg protein, although this cannot be demonstrated directly because biologically active levels of Wg are below the limits of antibody detection^{15–18}.

To investigate whether spatially graded Wg activity can generate different thresholds for activation of Dac and Dll, discs carrying multiple *Tkv** clones were double labelled (Fig. 5a). Ventrally located clones express Dll (arrowhead); more dorsally located clones express Dac but not Dll (arrow). Thus Dac can be induced at a greater distance from Wg than Dll, which suggests that Dac can be induced at levels of Wg activity that are not sufficient to induce Dll in the presence of a constant level of Dpp signalling activity. Consistent with this interpretation, Dac is induced in a larger region than Dll by a clone of Wg-expressing cells (Fig. 5c,d). The clone is located proximal to the normal Dac domain on the dorsal side of the disc (Fig. 5c). Clones expressing low levels of Wg (*act* $\gg$ Wg; Fig. 5d) induce Dll over a shorter range than clones expressing high levels of Wg (*C765/UAS* $\gg$ Wg; Fig. 2d). In both examples the position of the clone is equivalent. The *C765/UAS* $\gg$ Wg clones were examined 36 h after clone induction, whereas the *Act* $\gg$ Wg clones were examined 40 h after induction, so the apparent difference in range cannot be attributed to the time allowed for diffusion of Wg protein or expansion of the domains by growth. These observations support the proposal that the local concentration of Wg protein defines the spatial limits of Dll and Dac expression.

To compare the range over which Dpp can activate Dll and Dac, we examined discs carrying multiple clones in which the Wg pathway is constitutively activated by expression of the activated form of Armadillo¹⁵. Dorsally located clones express both Dac and Dll (Fig. 5b, right, arrowhead), whereas laterally located clones

express Dac but not Dll (arrows), suggesting that Dac can be activated at levels of Dpp that are not sufficient to induce Dll. Clones of cells expressing Dpp were induced to compare directly the range over which Dpp can induce Dll and Dac. The clones shown in Fig. 5f,g are located in a lateral region of the disc outside the endogenous Dll and Dac domains (as indicated in Fig. 5e). The pattern of Dac and Dll expression depends on the time allowed for Dpp expression after clone induction. After 24 h, Dac is expressed but Dll is not (Fig. 5f). By 40 h, Dll is induced and a gap appears in the centre of the Dac domain, producing an overlapping pattern of Dll and Dac expression (Fig. 5g, arrow). This pattern suggests that low levels of Dpp can induce Dac but not Dll, and that higher levels can induce Dll and repress Dac. The change in the expression patterns may reflect the time needed to accumulate higher levels of Dpp after induction of the clones.

Taken together, these results suggest that the spatial domains of Dac and Dll expression are defined by different threshold levels of both Wg and Dpp activities.

Wg and Dpp actively maintain Dac repression

Consistent with the proposal that increased levels of Dpp may repress Dac, we found that Dac can be expressed in *Mad* mutant clones located in the tarsal region of the leg disc (Fig. 6a). Similarly, Dac is derepressed in *dsh* mutant clones in the tarsal region (Fig. 6b). Derepression of Dac was observed in clones induced as late as mid third instar. These observations indicate that Wg and Dpp act directly to repress Dac in the centre of the disc. Thus Dpp and Wg define both the proximal and distal limits of the Dac domain, one by induction at a low level of combined activity, the other by repression at higher levels of combined activity. Further, we note that Dac repression is actively maintained by Wg and Dpp signalling long after Dac and Dll have been induced and are stably expressed in the absence of further signalling. These observations are reminiscent of the concentration-dependent effects of Activin, which can define both on and off thresholds for *Xbra* expression in *Xenopus* explants^{34,35}. However, although activation of Dac is non-reversible, as would be predicted by Gurdon's ratchet model for production of threshold responses to a concentration gradient³⁴, repression is reversible for some time. This reversibility is important for the generation of the final pattern (see below).

Importance of growth in P–D axis patterning

Although the Dll and Dac domains overlap considerably by the mid third instar, there is little or no overlap between them at earlier stages (Fig. 1). This suggests that the threshold levels of Wg and Dpp activity required to activate Dll are similar to those required to repress Dac in the centre of the disc in late second instar. As indicated above, Dac repression continues to require both Wg and Dpp signalling well into the third instar, whereas Dll expression depends only transiently on Wg and Dpp and, once activated, Dll does not require continued signalling. It is therefore possible that the Dll domain could expand beyond the effective range of the Wg and Dpp signals as the disc grows. If this is the case, cells near the edge of the Dll domain might be unable to maintain repression of Dac as they are displaced out of range of Wg and Dpp, leading to coexpression of the genes in a ring surrounding the Dac repression domain.

To test this possibility, we produced clones of cells capable of expressing the activated Dpp receptor *Tkv** under the control of a *Dll–Gal4* driver (*DllGal4/UAS* $\gg$ *Tkv**). The clones were induced in second instar, when the Dll and Dac domains are non-overlapping. Only clones in the Dll-expression domain can express *Tkv**, so expression of *Tkv** begins in the region where Dac is already repressed. This links expansion of the region in which the Dpp signal transduction pathway is activated to growth of the Dll domain, in a manner independent of the actual range of the Dpp signal. As the Dll domain expands, the clones carry constitutive

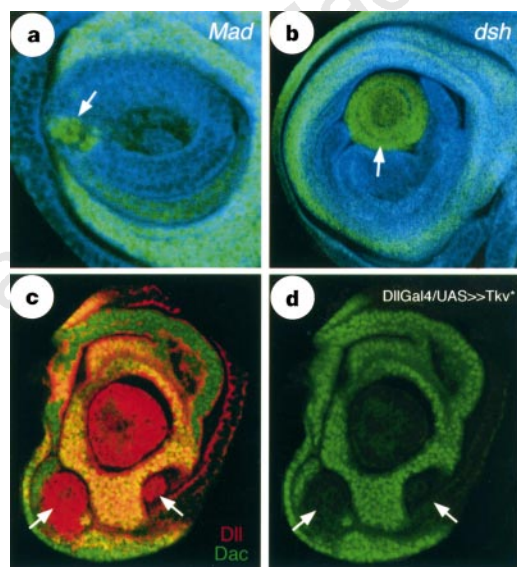


Figure 6 Repression of Dac by high levels of Wg and Dpp. **a**, *Mad* mutant clone in the centre of the leg disc (induced in mid third instar, ~84 h). **b**, *dsh* mutant clone induced in late second instar (~60 h). Clones are marked by the absence of β -gal (blue). Dac (green) is expressed in *Mad* or *dsh* mutant clones, suggesting that Dac is actively repressed by Wg and Dpp signals in the centre of the disc. Because Dac cannot be induced *de novo* in *Mad* or *dsh* mutant clones outside the endogenous domain, expression here must be due to alleviation of repression in cells already instructed to express Dac. Wg and Dpp may act together to induce repression of Dac in the centre of the disc. **c**, Leg disc double-labelled for Dac (green) and Dll (red). The disc carries clones of cells expressing the activated Dpp receptor under control of *Dll Gal4* (*DllGal4/UAS* $\gg$ *Tkv**, arrows). Clones were induced at 60 $\pm$ 12 h.

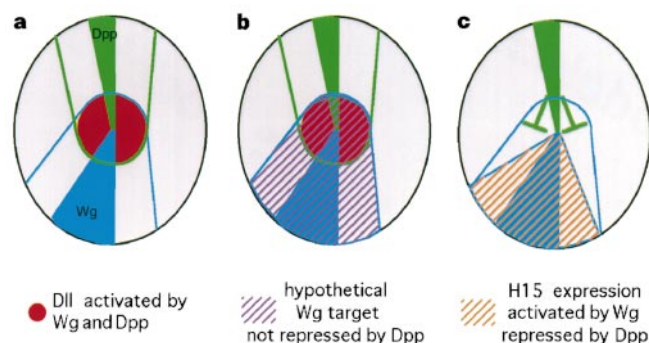


Figure 7 Roles of Wg and Dpp in the D-V and P-D axes. **a**, Wg and Dpp expression domains (blue and green, respectively). Blue and green lines represent the threshold levels of Wg and Dpp activities required for Dll induction (red). We propose that higher levels of one signal may compensate for lower levels of the second signal in the formation of circular expression domains in the centre of the disc. Note that these domains are specified in second instar, when

the disc is quite small. **b**, Expression pattern of a hypothetical target gene activated by Wg alone at the same threshold level as Dll (purple). Note that expression is both dorsal and ventral. **c**, Expression pattern of H15 (orange) relative to Wg and Dpp. H15 is activated by Wg and repressed by Dpp¹⁷. Dpp activity restricts H15 expression to the ventral side of the leg disc.

activation of the Dpp pathway into the region of overlap and maintain repression of Dac (Fig. 6c). Because Wg activity is also required, repression is observed on the ventral side of the disc where Wg is expressed. This observation shows how the Dll and Dac domains can make the transition from non-overlapping to significantly overlapping as a function of growth. The model is dependent on the finding that the signals initially used to specify Dll are only transiently required, whereas continued signalling is required to repress Dac. This is in contrast to the situation in the developing wing disc, where continuous signalling by both Wg and Dpp is required to maintain target gene expression along the A-P and D-V axes^{11,15,16}.

Wg and Dpp organize D-V and P-D axes

Two criteria have been applied to define Wg and Dpp as concentration-dependent morphogens in wing development: they act directly and at long distance to regulate expression of their target genes; and they specify distinct spatial domains of gene expression at different threshold activity levels^{11–13,15,16}. We have demonstrated that the combined action of Wg and Dpp in patterning the P-D axis fulfils both of these criteria. Our results suggest that Wg and Dpp act together in a manner dependent on the levels of both activities to specify discrete cell fates along the P-D axis of the leg. This graded activity is likely to reflect gradients of the proteins, but this cannot be demonstrated directly.

Wg and Dpp also pattern the D-V axis of the leg. The two signals act antagonistically to repress each other's expression and to specify dorsal and ventral fates^{17–19,29}. Wg specifies ventral cell fate and directs expression of its target gene *H15* in a stripe along the entire P-D axis of the leg on the ventral side¹⁷. As well as repressing Wg expression, Dpp blocks the ability of Wg to activate *H15*, suggesting that Wg and Dpp act antagonistically to define dorsal and ventral fates¹⁷. Thus at least one gene expressed in a specific D-V domain is regulated positively by one signal and negatively by the other. It remains to be seen whether other targets of Wg and Dpp are also regulated by antagonistic action of the two signals to define other D-V domains. In contrast, genes with expression dependent on the positive action of both signals will be expressed in domains defined by the overlap of the two signals in the centre of the disc (Fig. 7). These domains will form circles of different diameter that are centred on the distalmost tip of the leg, defining domains along the P-D axis (Fig. 7). Although induction of Dll and Dac in these domains occurs quite early, when the disc is small, the circular domain in which Dac is repressed provides a direct measure of the range of the combined signals at one activity threshold in mid third-

instar larvae. Thus Wg and Dpp simultaneously organize both D-V and P-D axes of the leg. A similar synergy between Wg and Dpp has been reported in the regulation of Ubx expression in the *Drosophila* embryonic midgut³⁶.

Our results show that the A-P, D-V and P-D patterning systems of the fly leg are coupled: Hedgehog regulates Dpp and Wg expression along the A-P axis, and Wg and Dpp work antagonistically to organize the D-V axis, and work cooperatively to organize the P-D axis. A-P, D-V and P-D patterning systems are also coupled in vertebrate limb development, although the mechanisms are rather different. Sonic hedgehog, Wnt7a, FGF4 and FGF8 signalling are interdependent in that they are required to maintain each other's activity to promote limb outgrowth^{37–40}. Although Wnt7a is involved in dorsal fate specification in the ectoderm⁴¹, there is no evidence for a comparable role for bone morphogenetic proteins (BMPs) in ventral fate specification. Perhaps the most striking difference is the use of fibroblast growth factor (FGF) signalling proteins for P-D outgrowth of the vertebrate limb. FGFs have been implicated in both the initial specification of the limb bud and in directing subsequent outgrowth^{38,40,42,43}. There is no evidence suggesting that FGFs are involved in P-D axis formation in the fly leg; rather, our results suggest that Wg and Dpp work together in the fly to perform functions analogous to those of FGFs in the vertebrate limb.

In conclusion, our findings highlight the importance of growth as a means of generating new positional information. We propose that spatial gradients in the activity of secreted signalling molecules and the time at which the signals are interpreted are both important determinants of P-D pattern in the developing leg. □

Methods

Clones of genetically marked cells expressing high levels of Dpp, Tkv* or Wg were made using the combined GAL4, flip-out system^{11,15}. Clones were induced in larvae of the following genotypes:

Dpp: *y w HSF1p1; UAS >> CD2y⁺ >> dpp; C765Gal4*;
Wg: *y w HSF1p1; UAS >> CD2 >> fluWg; C765Gal4*;
Tkv*: *y w HSF1p1; UAS >> CD2 >> Tkv**; *C765Gal4* and *y w HSF1p1; UAS >> CD2 >> Tkv*/DlGal4*;
activated armadillo: *y w HSF1p1; UAS >> CD2 >> fluΔArm; C765Gal4*.
Clones expressing low levels of Dpp were made using *tub >> f⁺ >> dpp¹⁰*. Clones expression low levels of Wg were made using *act >> y⁺ >> wg²¹*.

Lack-of-function mutant clones were induced in larvae of the following genotypes:

w HSF1p1; Mad^{1,2} FRT40A/armadillo-lacZ FRT40A (*Mad^{1,2}* is a strong hypomorphic allele^{12,28});

w HSF1p1; Mad^{1,2} pka^{89/18} FRT40A/armadillo-lacZ FRT40A (pka^{89/18} is a strong hypomorphic allele³¹)
w HSF1p1; tkv^{a12} FRT40A/armadillo-lacZ FRT40A (tkv^{a12} is a null allele¹¹)
w zw3^{D127} FRT18A/armadillo-lacZ FRT18A; HSF1p38/+ (on chromosome 2). seg-zw3^{D127} is a null allele⁴⁴;
dsh^{VA153} FRT18A/armadillo-lacZ FRT18A; HSF1p38/+. dsh^{VA153} is a null allele (Flybase).

Antibodies used were: mouse anti-Wg¹⁷; mouse anti-Dac²⁶; polyclonal mouse anti-Dll⁴⁵; rabbit anti-Dll⁴⁶; mouse anti-CD2 (Serotec); rabbit anti-β-gal (Cappel); mouse anti-flu (BabCo).

Received 14 February; accepted 19 May 1997.

1. Brook, W. J., Diaz-Benjumea, F. J. & Cohen, S. M. Organizing spatial pattern in limb development. *Annu. Rev. Cell Dev. Biol.* **12**, 161–180 (1996).
2. Lawrence, P. A. & Struhl, G. Morphogens, compartments and pattern: lessons from *Drosophila*? *Cell* **85**, 951–961 (1996).
3. Diaz-Benjumea, F. J. & Cohen, S. M. Interaction between dorsal and ventral cells in the imaginal disc directs wing development in *Drosophila*. *Cell* **75**, 741–752 (1993).
4. Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by hedgehog protein. *Nature* **368**, 208–214 (1994).
5. Tabata, T. & Kornberg, T. Hedgehog is a signalling protein with a key role in patterning *Drosophila* imaginal discs. *Cell* **76**, 89–102 (1994).
6. Williams, J. A., Paddock, S. W., Vorwerk, K. & Carroll, S. B. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. *Nature* **368**, 299–305 (1994).
7. Irvine, K. & Wieschaus, E. *fringe*, a boundary specific signalling molecule, mediates interactions between dorsal and ventral cells during *Drosophila* wing development. *Cell* **79**, 595–606 (1994).
8. Kim, J., Irvine, K. D. & Carroll, S. B. Cell recognition, signal induction and symmetrical gene activation at the dorsal/ventral boundary of the developing *Drosophila* wing. *Cell* **82**, 795–802 (1995).
9. Diaz-Benjumea, F. J. & Cohen, S. M. Serrate signals through Notch to establish a Wingless-dependent organizer at the dorsal/ventral compartment boundary of the *Drosophila* wing. *Development* **121**, 4215–4225 (1995).
10. Zecca, M., Basler, K. & Struhl, G. Sequential organizing activities of *engrailed*, *hedgehog* and *decapentaplegic* in the *Drosophila* wing. *Development* **121**, 2265–2278 (1995).
11. Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a Dpp morphogen gradient. *Cell* **85**, 357–368 (1996).
12. Lecuit, T. *et al.* Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. *Nature* **381**, 387–393 (1996).
13. Singer, M. A., Penton, A., Twombly, V., Hoffmann, F. M. & Gelbart, W. M. Signaling through both type I Dpp receptors is required for anterior-posterior patterning of the entire *Drosophila* wing. *Development* **124**, 79–89 (1997).
14. Neumann, C. J. & Cohen, S. M. A hierarchy of cross-regulation involving *Notch*, *wingless*, *vestigal* and *cut* organizes the dorsal/ventral axis of the *Drosophila* wing. *Development* **122**, 3477–3485 (1996).
15. Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a Wingless morphogen gradient. *Cell* **87**, 833–844 (1996).
16. Neumann, C. J. & Cohen, S. M. Long-range action of Wingless organizes the dorsal–ventral axis of the *Drosophila* wing. *Development* **124**, 871–880 (1997).
17. Brook, W. J. & Cohen, S. M. Antagonistic interactions between Wingless and Decapentaplegic responsible for dorsal–ventral pattern in the *Drosophila* leg. *Science* **273**, 1371–1377 (1996).
18. Jiang, J. & Struhl, G. Complementary and mutually exclusive activities of Decapentaplegic and Wingless organize axial pattern during *Drosophila* limb development. *Cell* **86**, 401–409 (1996).
19. Penton, A. & Hoffmann, F. M. Decapentaplegic restricts the domain of *wingless* during *Drosophila* limb patterning. *Nature* **382**, 162–165 (1996).
20. Cousens, P., Tate, M. & Martinez Arias, A. A *wingless*-dependent polar coordinate system in the imaginal discs of *Drosophila*. *Science* **259**, 484–489 (1993).

21. Struhl, G. & Basler, K. Organizing activity of *wingless* protein in *Drosophila*. *Cell* **72**, 527–540 (1993).
22. Diaz-Benjumea, F. J., Cohen, B. & Cohen, S. M. Cell interactions between compartments establishes the proximal–distal axis of *Drosophila* limbs. *Nature* **372**, 175–179 (1994).
23. Held, L. J., Heup, M. A., Sappington, J. M. & Peters, S. D. Interactions of *decapentaplegic*, *wingless* and *Distal-less* in the *Drosophila* leg. *Wilhelm Roux Arch. Dev. Biol.* **203**, 310–319 (1994).
24. Campbell, G., Weaver, T. & Tomlinson, A. Axis specification in the developing *Drosophila* appendage: the role of *wingless*, *decapentaplegic*, and the homeobox gene *aristaleless*. *Cell* **74**, 1113–1123 (1993).
25. Cohen, S. M., Brönnner, G., Küttner, F., Jürgens, G. & Jäckle, H. *Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*. *Nature* **338**, 432–434 (1989).
26. Mardon, G., Solomon, N. M. & Rubin, G. M. *dachshund* encodes a nuclear protein required for normal eye and leg development in *Drosophila*. *Development* **120**, 3473–3486 (1994).
27. Hoodless, P. A. *et al.* MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* **85**, 489–500 (1996).
28. Wiersdorff, V., Lecuit, T., Cohen, S. M. & Mlodzik, M. *Mad* acts downstream of the Dpp receptors and reveals differential requirement for *dpp* signaling between initiation and propagation of morphogenesis in the *Drosophila* eye. *Development* **122**, 2153–2162 (1996).
29. Theisen, H., Haerry, T. E., O'Connor, M. B. & Marsh, J. L. Developmental territories created by mutual antagonism between Wingless and Decapentaplegic. *Development* **122**, 3939–3948 (1996).
30. Jiang, J. & Struhl, G. Protein kinase A and Hedgehog signaling in *Drosophila* limb development. *Cell* **80**, 563–572 (1995).
31. Lepage, T., Cohen, S. M., Diaz-Benjumea, F. J. & Parkhurst, S. M. Signal transduction by cAMP-dependent protein kinase A in *Drosophila* limb patterning. *Nature* **373**, 711–715 (1995).
32. Li, W., Ohlmeyer, J. T., Lane, M. E. & Kalderon, D. Function of protein kinase A in Hedgehog signal transduction and *Drosophila* imaginal disc development. *Cell* **80**, 553–562 (1995).
33. Pan, D. J. & Rubin, G. Protein kinase and hedgehog act antagonistically in regulating *decapentaplegic* transcription in *Drosophila* imaginal discs. *Cell* **80**, 543–562 (1995).
34. Gurdon, J. B., Mitchell, A. & Mahoney, D. Direct and continuous assessment by cells of their position in a morphogen gradient. *Nature* **376**, 520–521 (1995).
35. Gurdon, J. B., Harger, P., Mitchell, A. & Lemaire, P. Activin signalling and response to a morphogen gradient. *Nature* **371**, 487–492 (1994).
36. Riese, J. *et al.* LEF-1, a nuclear factor coordinating signal inputs from *wingless* and *decapentaplegic*. *Cell* **88**, 777–787 (1997).
37. Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. *Sonic hedgehog* and *Fgf-4* act through a signalling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993–1003 (1994).
38. Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning of the vertebrate limb. *Nature* **371**, 609–612 (1994).
39. Yang, Y. & Niswander, L. Interaction between the signalling molecules WNT7a and SHH during vertebrate limb development: dorsal signals regulate anteroposterior patterning. *Cell* **80**, 939–947 (1995).
40. Crossley, P. H., Minowada, G., MacArthur, C. A. & Martin, G. R. Roles for FGF-8 in the induction, initiation and maintenance of chick limb development. *Cell* **84**, 127–136 (1996).
41. Parr, B. A. & McMahon, A. Dorsalizing signal *Wnt-7a* required for normal polarity of A/P and D/V axes on mouse limb. *Nature* **374**, 350–353 (1995).
42. Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* **75**, 579–587 (1993).
43. Cohn, M. J., Izpisua-Belmonte, J. C., Abud, H., Heath, J. K. & Tickle, C. Fibroblast growth factors induce additional limb development from the flank of chick embryos. *Cell* **80**, 739–746 (1995).
44. Ruel, L., Pantescio, V., Lutz, Y., Simpson, P. & Bourouis, M. Functional significance of a family of protein kinases encoded at the *shaggy* locus in *Drosophila*. *EMBO J.* **12**, 1657–1669 (1993).
45. Vachon, G. *et al.* Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. *Cell* **71**, 437–450 (1992).
46. Panganiban, G., Sebring, A., Nagy, L. & Carroll, S. B. The development of crustacean limbs and the evolution of arthropods. *Science* **270**, 1363–1366 (1995).

Acknowledgements. We thank G. Struhl, K. Basler and K. Cadigan for fly strains; G. Panganiban and G. Mardon for antibodies; and M. Averof, W. Brook, J. Royet and K. Weigmann for suggestions on the manuscript.

Correspondence and requests for materials should be addressed to S.C. (e-mail: scohen@embl-heidelberg.de).

A dark cluster of galaxies at redshift $z = 1$

M. Hattori^{*†‡}, Y. Ikebe[†], I. Asaoka^{*†}, T. Takeshima^{†§},
H. Böhringer^{*}, T. Mihara[†], D. M. Neumann^{*},
S. Schindler^{*||}, T. Tsuru[¶] & T. Tamura[#]

^{*} Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, 85740 Garching, Germany

[†] The Institute of Physical and Chemical Research (Riken), Hirosawa Wako, Saitama 351-01, Japan

[‡] Astronomical Institute, Tohoku University, Aoba Aramaki, Sendai 980, Japan
[§] Laboratory for High Energy Astrophysics, NASA/GSFC (USRA), Code 662, Greenbelt, Maryland 20771, USA

^{||} Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85740 Garching, Germany

[¶] Department of Physics, Kyoto University, Kyoto, Sakyo-ku, Japan

[#] Department of Physics, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

The abundance of metals in the hot, gaseous X-ray haloes of galaxy clusters depends crucially on the evolution of the constituent galaxies and their associated stellar populations. The metal abundances in X-ray clusters at high redshifts should therefore provide important insights into the nature and epoch of galaxy formation. Here we report the detection of an extended X-ray source in the direction of the lensed quasi-stellar object MG2016 + 112 (refs 1, 2). Although deep optical searches have failed to reveal a galaxy cluster at the lens position^{3,4}, the X-ray emission is consistent with thermal bremsstrahlung radiation from a hot metal-rich, diffuse gaseous halo, as observed in nearby galaxy clusters. This is the most distant galaxy cluster discovered in X-rays so far. Furthermore, the mass of the cluster derived from this emission is consistent with that implied by lensing models of the system⁵. Given that the cluster apparently comprises few galaxies, yet contains a large amount of iron, a new type of astronomical object is implied by our results. A revision of theoretical models of the metal enrichment process in galaxy clusters may therefore be required.

A deep observation of MG2016 + 112 with the ASCA satellite X-ray observatory was performed from 24 to 26 October 1994, with an effective exposure time of 85 ks. A new X-ray source (referred to as AX J2019 + 1127) was discovered in the direction of the lens system with a significance of 10σ . The position of the source is coincident with the location of all components of MG2016 + 112 taking into

account the positional accuracy of 1 arcmin for both telescopes equipped with Gas Imaging Spectrometers (GISs).

The source could originate from either the high-redshift quasar at redshift 3.27 or from a cluster centred on a giant elliptical galaxy, named galaxy D, at redshift 1.01 which is producing the lensing effect^{3,5,6}. A clue to resolving this question comes from X-ray spectroscopy. Figure 1 shows the X-ray spectrum of AX J2019 + 1127 obtained by ASCA. A single continuum model—that is, a bremsstrahlung with three free parameters (temperature, strength and absorption column density)—gives a statistically acceptable fit to the GIS and SIS (solid imaging spectrometer) spectrum simultaneously ($\chi^2 = 56.66$ with 54 degrees of freedom d.o.f.); however, an emission line feature is clearly seen at 3.5 keV. Inclusion of an emission line in a bremsstrahlung model reduces χ^2 by 9.4 which, for two additional free parameters (line energy and strength), is significant at the 99.1% level. This shows that the existence of the emission line is statistically significant. The best-fit line energy is $3.49^{+0.11(+0.15)}_{-0.09(-0.13)}$ keV where the first errors are the 90% confidence level statistical errors and the errors in parentheses represent the $\pm 1\%$ systematic error in the ASCA energy scale calibration. (The statistical errors are estimated by treating all the parameters as free except the one of interest, in this case, the line energy.) This implies a redshift of $z = 0.92^{+0.05(+0.07)}_{-0.06(-0.08)}$ if the line is due to Fe XXV or $z = 1.00^{+0.05(+0.07)}_{-0.06(-0.08)}$ if due to Fe XXVI. If the source is a cluster of galaxies, the line is a combination of Fe XXVI and Fe XXV. Thus the line matches well with the redshift of galaxy D. Assuming that the source is a cluster, the spectrum can be fitted with a Raymond–Smith plasma model⁷. The inclusion of a non-zero abundance reduces χ^2 by 10.01 and is significant at the 99.9% level (3σ). The fitted redshift is $z = 0.94^{+0.05(+0.07)}_{-0.06(-0.08)}$. The fitted intra-cluster gas temperature is $kT = 8.6^{+4.2}_{-3.0}$ keV (errors are 1σ statistical errors). We find an X-ray luminosity of $L_X = (8.4^{+2.4}_{-1.7}) \times 10^{44} h_{50}^{-2} \text{ erg s}^{-1}$ (2–10 keV; we assume in this Letter that $H_0 = 50 h_{50} \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 0.5$) and an iron abundance of $Z = 1.7^{+1.25}_{-0.74} Z_\odot$ (where $Z_\odot$ is the solar value) with an acceptable fit yielding χ^2 with 52 degrees of freedom where the solar iron abundance is 4.67×10^{-5} (ref. 8). The best-fit absorption hydrogen column density of $N_H = (0.18^{+0.15}_{-0.10}) \times 10^{22} \text{ cm}^{-2}$ is consistent with the Galactic value of $N_H = 0.15 \times 10^{22} \text{ cm}^{-2}$ (ref. 9). The correlation of X-ray gas temperature and X-ray luminosity for nearby clusters^{10,11} and the observed X-ray luminosity of AX J2019 + 1127 gives a temperature of ~ 7.5 keV, in good agreement with the observed value. These results are consistent with the idea that there is a cluster of galaxies including the galaxy D as a member galaxy. A single power-law model also gives an acceptable fit to the continuum ($\chi^2/\text{d.o.f.} = 58.11/54 = 1.08$) with a fitted

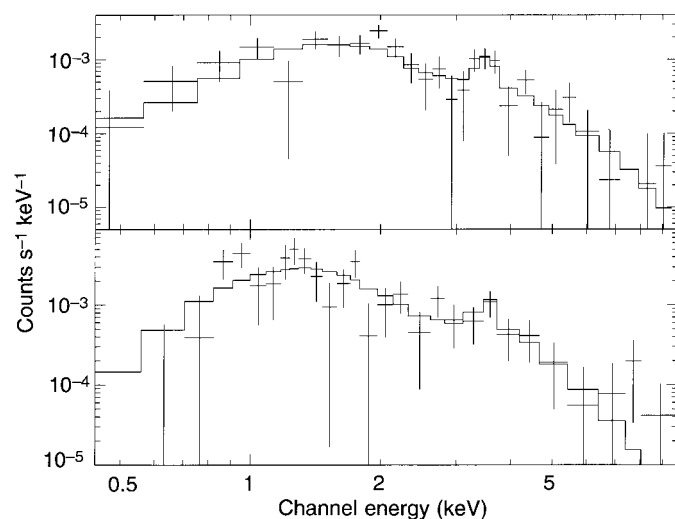


Figure 1 X-ray spectrum of AX J2019 + 1127 obtained by ASCA. The ASCA observatory has four identical telescopes covering the energy range 0.3–10 keV (ref. 22). They are equipped with two gas imaging spectrometers (GIS 2 and 3) and two solid-state imaging spectrometers (SIS 0 and 1) at the focal plane. The background-subtracted source spectrum obtained by two GISs (top panel) and by two SISs (bottom panel) with 1σ Poisson errors are shown. The solid lines are the best-fit Raymond–Smith models obtained by simultaneous fitting of the GIS and SIS spectra. The background is taken from blank field observations during the test phase of the ASCA mission. The SIS data were taken with two CCD bright mode observations with an effective exposure time of 22 ks.

photon index of 1.8 ± 0.2 . The derived slope is consistent with that found for active galaxies¹². If the object is an active galaxy, the detected iron line should be fluorescent Fe and this implies a redshift of $z = 0.83^{+0.03(+0.05)}_{-0.03(-0.05)}$. However, this is clearly inconsistent with the redshift of the quasi-stellar object (QSO) and there is no known object, at this redshift, inside the error circle.

Another crucial test for the cluster origin of the X-ray emission would be to resolve the source extent. This was performed with a deep observation with the High Resolution Imager (HRI) on the Rosat satellite. We find a diffuse source with 76 ± 24 source counts, $(1.4 \pm 0.4) \times 10^{-3}$ counts s^{-1} within a circle of radius 1 arcmin. Figure 2 shows the source image. The peak detection significance of the source is 4.2σ . The maximum of the X-ray emission is consistent with the position of the cD galaxy D within the HRI pointing accuracy. The HRI count rate is consistent with, but slightly less than, that expected from the ASCA observation, that is, $\sim 2.5 \times 10^{-3}$ counts s^{-1} . We confirm that there is no other source above the 3σ significance detection limit in the circle of radius 3 arcmin centred on MG2016 + 112 by applying a standard maximum-likelihood analysis of the source content of the HRI field. We therefore conclude that the source detected by the Rosat HRI is identical to AX J2019 + 1127. Figure 3 shows that the X-ray surface brightness distribution of the source deviates significantly from the profile of the Rosat HRI point-spread function and is more extended. The significance of the deviation is estimated to be $\sim 4\sigma$. From this we conclude that the source is extended. As the source counts within a concentric circle increases monotonically up to 1 arcmin, the source radius is estimated to be 1 arcmin ($500h_{50}^{-1}$ kpc for $z = 1$). The optical quasar consists of two almost equally bright images separated by 3.4 arcsec. For the point-spread function of the HRI with full-width at half-maximum of 5 arcsec, the X-ray surface brightness for two X-ray images at 3.4-arcsec separation would not be significantly different from the point

source result. Therefore, the extent of AX J2019 + 1127 cannot be explained if the quasar is the source. The low central surface brightness implies that the possible X-ray flux of the QSO contributes at most 10% to the observed value. This is consistent with the flux of the two bright QSO images as expected from their optical luminosity¹, based on an optical to X-ray correlation¹³, that is, 1.2×10^{-14} erg $cm^{-2} s^{-1}$ in the band 2–10 keV and an HRI count rate of 0.16×10^{-3} counts s^{-1} .

The above results imply that AX J2019 + 1127 is a galaxy cluster in which galaxy D is the central dominant galaxy. This is the most distant galaxy cluster discovered in X-rays so far. Because the X-ray emission provides the best evidence that clusters are gravitationally bound entities (unless a large number of measured galaxy redshifts are available), this is also the most distant clearly confirmed galaxy cluster. The redshift of the next furthest known X-ray cluster is MS1054.5 – 0321 at $z = 0.826$ (refs 14, 15). We can now further test if the observations are consistent with the mass distribution model needed for the lensing effect (for example, model II in ref. 5). Assuming that the cluster is isothermal and in hydrostatic equilibrium, one can obtain a rough estimate of the cluster mass. The best-fit results yields a central electron density of $1.7 \times 10^{-2} h_{50}^{0.5} cm^{-3}$, a gas mass of $2.4 \times 10^{13} h_{50}^{-2.5} M_{\odot}$ and a gravitational mass of $3.6 \times 10^{14} (kT/8.6 \text{ keV}) h_{50}^{-1} M_{\odot}$ within a sphere of radius of $500h_{50}^{-1}$ kpc in the cluster (here $M_{\odot}$ is the solar mass). The gas mass fraction of $\sim 7h_{50}^{-1.5}\%$ of the total mass is typical for the central regions of nearby clusters¹⁶ within the measurement errors.

These results are consistent with the mass model for the lens proposed by ref. 5. It implies that we have indeed discovered the object responsible for the lensing effect. Another important result of this discovery is the high iron abundance detected in the intra-cluster medium of AX J2019 + 1127. Although the errors on the abundance are quite large, the lower limit on the iron content in this cluster is at least as high as that for nearby clusters of galaxies¹⁷.

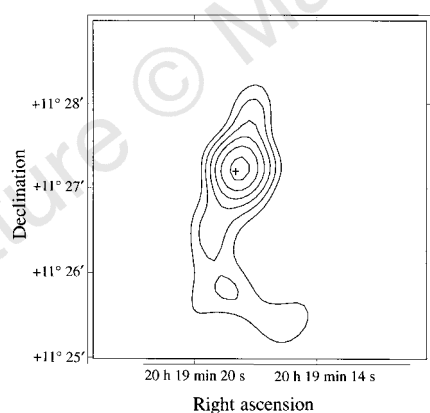


Figure 2 X-ray image of AX J2019 + 1127 obtained by the Rosat HRI. A deep Rosat HRI²³ observation was performed to resolve the possible extent of the source (right ascension and declination are equinox 2000.0). The lens system was observed from 15 to 20 November 1995, from 19 to 28 April 1996 and from 22 October to 13 November 1996 with a total exposure time of 79 ks, yielding an effective exposure of 53.7 ks after removal of observing times with high background flux. The accuracy of the HRI pointing is calibrated with two point sources coincident with stars in the Palomar Sky Survey observed in the field. The accuracy of the HRI pointing is better than 9 arcsec. The contours show the significance levels with a linear spacing of 0.5σ from the minimum level of 1.5σ . The peak significance is 4.2σ . The image is smoothed with a gaussian filter with $\sigma = 15$ arcsec. The cross shows the position of galaxy D.

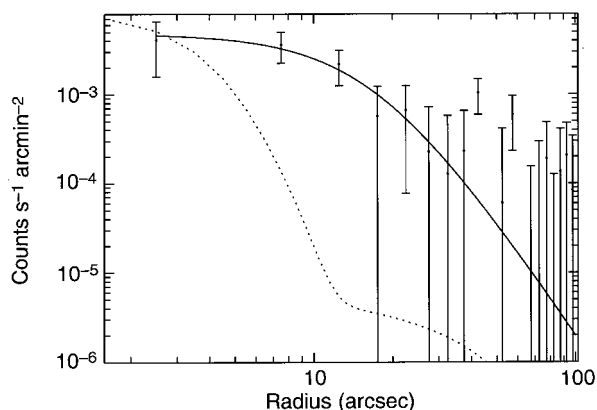


Figure 3 The background-subtracted X-ray surface brightness profile of AX J2019 + 1127 obtained by the Rosat HRI (data points). The profile of the Rosat HRI point-spread function at the same off-axis angle as the source position is also shown (dashed line). The central surface brightness of the point-spread function is set for a point source at the same central surface brightness as the source. Superimposed on the data is the best-fit isothermal β model (solid line) using a core radius of $17'' \approx 150h_{50}^{-1}$ kpc and $\beta = 0.9$, values which are typical for X-ray profiles of nearby rich clusters²⁴. The central surface brightness of the model is $\sim 3.3 \times 10^{-3}$ counts $s^{-1} arcmin^{-2}$.

Currently the most distant galaxy cluster in which an Fe K-line has been detected is at a redshift of $z = 0.54$ (ref. 18). Therefore, AXJ2019 + 1127 is now the highest redshift cluster for which the existence of a metal-enriched intra-cluster medium has been confirmed. The detection of the large iron content at high redshift sets a new limit for the epoch of the enrichment. This favours models where early star burst phases in galaxies are responsible for the metal enrichment of the intra-cluster medium^{19,20}.

More surprisingly, a high iron abundance has been discovered in this 'dark cluster', which is very poor in its galaxy content according to the deep optical searches^{3,4}. The main optical light source from this cluster is galaxy D which has a blue luminosity of $L_B = 1.1 \times 10^{11} h_{50}^{-2} L_\odot$ (where $L_\odot$ is the solar luminosity). Another possible member galaxy is galaxy C (ref. 1) which, however, can contribute only a fraction of the optical luminosity of galaxy D. Therefore, the mass-to-light ratio of the cluster is $ML_B \approx 3,300(kT/8.6 \text{ keV})^{-1} h_{50} M_\odot/L_\odot$. Because, according to our current understanding, the metal enrichment originates in the stars of the cluster galaxies, it is very puzzling that we detect such a high iron abundance in this 'dark cluster'. Therefore, either AXJ2019 + 1127 is a very new and enigmatic object or it has more cluster members at fainter magnitudes that have escaped the deep searches conducted so far. Recently, it has been reported²¹ that there are also some low-redshift clusters that are optically dark but bright in X-rays. Combined with our results, this may indicate that there exist many such objects in our Universe which have eluded optical identification. In any case, AXJ2019 + 1127 provides us with a new means for testing cosmological theory and ought to be well studied in various wavelength bands. □

Received 6 December 1996; accepted 21 May 1997.

1. Lawrence, C. R. *et al.* Discovery of a new gravitational lens system. *Science* **223**, 46–49 (1984).
2. Lawrence, C. R. in *Astrophysical Applications of Gravitational Lensing* (eds Kochanek, C. S. & Hewitt, J. N.) 299–304 (Springer, Berlin, 1995).
3. Schneider, D. P. *et al.* Deep optical and radio observations of the gravitational lens system 2016+112. *Astrophys. J.* **294**, 66–69 (1985).
4. Langston, G. I., Fischer, J. & Aspin, C. Infrared K-band observations of the gravitational lens 2016+112. *Astron. J.* **102**, 1253–1257 (1991).
5. Narasimha, D., Subramanian, K. & Chitre, S. M. The gravitational lens system 2016 + 112 revised. *Astrophys. J.* **315**, 434–439 (1987).
6. Schneider, D. P. *et al.* The third image, the redshift of the lens, and new components of the gravitational lens 2016 + 112. *Astron. J.* **91**, 991–997 (1986).
7. Raymond, J. C. & Smith, B. W. Soft X-ray spectrum of a hot plasma. *Astrophys. J. Suppl. Ser.* **35**, 419–439 (1977).
8. Anders, E. & Grevesse, N. Abundances of the elements—meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
9. Dickey, J. M. & Lockman, F. J. HI in the Galaxy. *Annu. Rev. Astron. Astrophys.* **28**, 215–261 (1990).
10. David, L., Slyz, A., Jones, C., Forman, W. & Vrtilsek, S. D. A catalog of intracluster gas temperatures. *Astrophys. J.* **412**, 479–488 (1993).
11. Tsuru, T. *et al.* in *The 11th Int. Colloq. on UV and X-ray Spectroscopy of Astrophysical and Laboratory Plasmas* (eds Watanabe, T. & Yamashita, K.) 375–378 (Universal Academy, Tokyo, 1996).
12. Williams, O. R. *et al.* The X-ray spectra of high-luminosity active galactic nuclei observed by GINGA. *Astrophys. J.* **389**, 157–178 (1992).
13. La Franca, F., Franceschini, A., Cristiani, S. & Vio, R. On the relationship between the optical and X-ray luminosities of quasars. *Astron. Astrophys.* **299**, 19–24 (1995).
14. Gioia, I. M. & Luppino, G. A. The EMSS Catalog of X-ray selected clusters of galaxies. I. An atlas of CCD images of 41 distant clusters. *Astrophys. J. Suppl. Ser.* **94**, 583–614 (1994).
15. Luppino, G. A. & Gioia, I. M. Constraints on cold dark matter theories from observations of massive X-ray-luminous clusters of galaxies at high redshift. *Astrophys. J.* **445**, L77–L80 (1995).
16. Briel, U. G., Henry, J. P. & Böhringer, H. Observation of the Coma cluster of galaxies with ROSAT during the All-Sky Survey. *Astron. Astrophys.* **259**, L31–L34 (1992).
17. Ohashi, T. in *Proc. 17th Texas Symp. on Relativistic Astrophysics and Cosmology* (eds Böhringer, H., Morfill, G. E. & Trümper, J. E.) 217–220 (New York Acad. Sci., New York, 1995).
18. Donahue, M. Temperature and metallicity of a massive X-ray cluster at redshift 0.5. *Astrophys. J.* **468**, 79–85 (1996).
19. Arnaud, M., Rothenflug, R., Boullade, O., Vigroux, L. & Vangioni-Flam, E. Some constraints on the origin of the iron enriched intra-cluster medium. *Astron. Astrophys.* **254**, 49–64 (1992).
20. Hattori, M. & Terasawa, N. Metal enrichment of the intracluster medium from protogalaxies and the gamma-ray bursts. *Astrophys. J.* **406**, L55–L58 (1993).
21. Ebeling, H., Mendes de Oliveira, C. & White, D. A. A2572 and HCG94—galaxy clusters but not as we know them: an X-ray case study of optical misclassifications. *Mon. Not. R. Astron. Soc.* **277**, 1006–1032 (1995).
22. Tanaka, Y., Inoue, H. & Holt, S. S. The X-ray astronomy satellite ASCA. *Publ. Astron. Soc. Jpn* **46**, L37–L41 (1994).
23. Trümper, J. E. ROSAT. *Phys. Scripta* **T7**, 209–215 (1984).
24. Sarazin, C. L. X-ray emissions from clusters of galaxies. *Rev. Mod. Phys.* **58**, 1–116 (1986).

Acknowledgements. We thank T. Sonobe, T. Miyaji, C. R. Lawrence, Y. Tanaka and M. Matsuoka for discussions. M.H. was supported in part by the post-doctoral program of the Max-Planck Gesellschaft and Yamada Science Foundation; Y.L. was supported by the Special Researchers' Basic Science Program of the Riken. H.B. and S.S. thank the Verbundforschung for support.

Correspondence should be addressed to M.H. (e-mail: hattori@astr.tohoku.ac.jp).

Properties of ideal composite knots

Vsevolod Katritch*, Wilma K. Olson*, Piotr Pieranski†, Jacques Dubochet‡ & Andrzej Stasiak‡

* Department of Chemistry, Rutgers the State University of New Jersey, New Brunswick, New Jersey 08903, USA

† Institute of Physics, Poznan University of Technology, 60-965 Poznan and Institute of Molecular Physics, 60-159 Poznan, Poland

‡ Laboratoire d'Analyse Ultrastructurale, Bâtiment de Biologie, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland

The shortest tube of constant diameter that can form a given knot represents the 'ideal' form of the knot^{1,2}. Ideal knots provide an irreducible representation of the knot, and they have some intriguing mathematical and physical features, including a direct correspondence with the time-averaged shapes of knotted DNA molecules in solution^{1,2}. Here we describe the properties of ideal forms of composite knots—knots obtained by the sequential tying of two or more independent knots (called factor knots) on the same string. We find that the writhe (related to the handedness of crossing points) of composite knots is the sum of that of the ideal forms of the factor knots. By comparing ideal composite knots with simulated configurations of knotted, thermally fluctuating DNA, we conclude that the additivity of writhe applies also to randomly distorted configurations of composite knots and their corresponding factor knots. We show that composite knots with several factor knots may possess distinct structural isomers that can be interconverted only by loosening the knot.

The ideal representation of a knot corresponds to the shortest axial path of an impenetrable tube of uniform thickness forming a given knot type¹. In other words, the ideal form of a knot is the spatial arrangement of the knotted tube with uniform thickness for which the length-to-diameter ratio is the smallest. These configurations can be obtained in computer simulations that maximize the thickness of knotted tubes while maintaining a fixed contour length. In the case of prime knots, computer runs end (in most cases), in the same final configuration, irrespective of the starting configuration of a given knot¹. For composite knots, especially those involving asymmetric factor knots, the simulations frequently end in local minima. To find the global minimum of the length-to-diameter ratio, it is necessary to carry out many independent simulations, testing different ways of combining two factor knots. For example, in searching for the ideal configuration of a composite knot composed of 6_1 and 3_1 factor knots, we created initial configurations where the symmetric 3_1 knot was attached to different loops of the asymmetric 6_1 knot. (The designations of knots follow the standard notation; the first number indicates the minimal number of crossings for a given knot and the subscript denotes the position of the knot among all other knots with the same crossing number³.) When both factor knots are symmetric, for example, 3_1 or 4_1 , all loops are equivalent and there is no need to test many initial configurations. For different joinings of factor knots, the simulations frequently end in visually distinct configurations with similar length-to-diameter ratios, whereas simulations started from a given type of joining usually result in the same kinds of configurations with nearly identical length to diameter ratios. Although we are not sure if we have found the configurations with the smallest possible length-to-diameter ratios for the composite knots analysed, for the purpose of this Letter we designate as ideal the best configurations of the different knots that we have obtained in computer simulations. Figure 1 thus shows the ideal configurations of 16 composite knots where the factor knots range from trefoils up to six-noded knots.

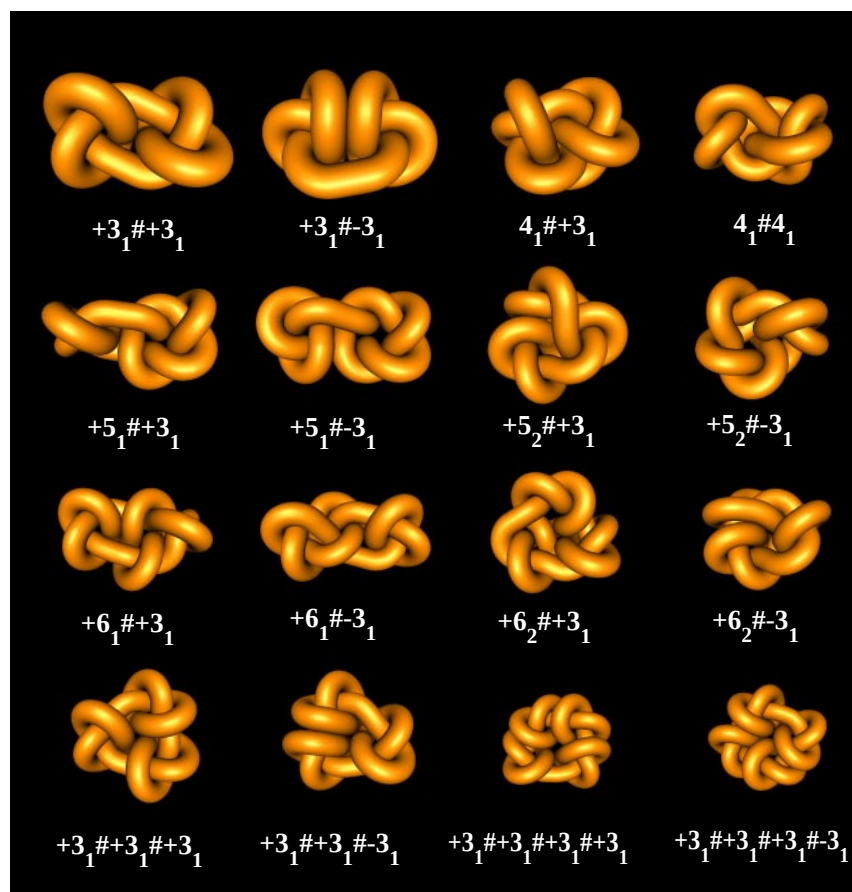


Figure 1 Ideal geometrical representations of several kinds of composite knots as viewed down their longest principal axes of inertia. (Three-dimensional models of these knots in VRML format are available at <http://rutchem.rutgers.edu/~seva/composite.html>). The designations of the different composite knot types correspond to the factor knots that comprise them. For example, the designation $+3_1\#4_1$ denotes a composite knot built from a right-handed trefoil and an achiral four-noded knot. We use the designations for factor knots given in standard tables of knots³. To indicate the chirality of a factor knot, we introduce a + or – sign, respectively, for right- or left-handed types. The absence of a sign indicates that the given factor knot is achiral. All ideal knot configurations shown here have the same axial length. As a result, the ratio between the contour lengths and diameters of the tubes making up the ideal knots increases progressively as the configurations become more complicated¹. The optimized configurations are the result of a joint effort of two simulation programs based on two independent algorithms. The first one⁹—simpler and faster, but less precise—simulates a mechanical process in which a completely flexible, frictionless tube, forming a knot of a given type, is uniformly inflated. Configurations found with this procedure are used as the starting configurations for the Metropolis–Monte Carlo simulations performed with the second program¹.

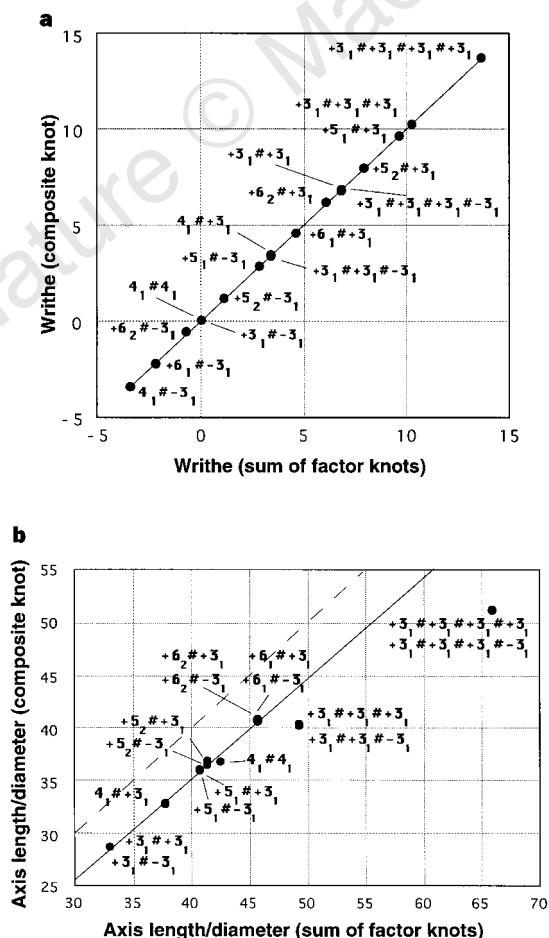


Figure 2 Additivity of the writhe and subadditivity of the relative lengths of ideal representations of composite knots. **a**, The 1:1 correspondence between the writhes of the ideal representations of different composite knots and the sums of the writhes of the respective factor knots in their ideal configurations. **b**, The relative length of the composite knots is always smaller (by nearly a constant value) than the sum of the relative lengths of the factor knots constituting a given composite knot. The dashed line denotes a hypothetical situation where the relative lengths of the composite knots correspond to the sums of the relative lengths of the factor knots. Composite knots with three and four factor knots show correspondingly bigger deficits of the relative length.

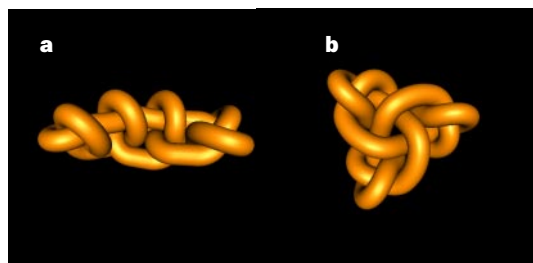


Figure 3 Linear (a) and star-shaped (b) configurations of composite $+3_1\#3_1\#3_1$ knot.

To characterize the geometry of the ideal representations of the knots, it is informative to calculate their writhe, an accepted measure of the handedness of a curve in space⁴. One way to measure the writhe of a given knot is to observe the knot from many uniformly spread out directions on the unit sphere and to count the average number of right- and left-handed crossings per projection⁴. Right-handed crossings are scored as +1 and left-handed ones as -1. For example, the writhe of an ideal right-handed trefoil knot is close to +3.4 (ref. 1), a number which indicates that on average there are 3.4 more right-handed than left-handed crossings found when observing the knot from random directions. The data presented in Fig. 2a demonstrate that the writhe of the ideal configuration of a composite knot is equal to the sum of the writhe values of the ideal configurations of the constituent factor knots. The observed additivity of writhe is somewhat surprising because in general the writhe of a space curve is not additive. That is, if one arbitrarily joins the two complicated curves in space by removing parts of the original curves and replacing them with other segments, the writhe of the new curve is usually different from the sum of the writhes of the two original curves⁵. But in the ideal configurations of composite knots, even when the factor knots change slightly from the shapes that they assume as independent knots, these changes are self-compensatory and the writhe remains the same.

We have checked whether the mean writhe of the pool of thermally distorted polymer chains closed into different composite knots attains the same value as the writhe of the corresponding ideal composite knot. Taking advantage of our experience in modelling configurational ensembles of DNA knots under equilibrium conditions in solution¹, we have used a Metropolis–Monte Carlo procedure to simulate millions of thermally distorted configurations of nicked DNA molecules closed in a given composite knot. As in our previous study, we have simulated knotted molecules of 5,400 base pairs, corresponding to 18 Kuhn statistical segments (a size range typical of bacterial plasmid DNA). During the simulation the DNA chains are permitted to relax any torsional stress as would be the case for DNA containing interruptions in one of the strands of the double helix. Table 1 shows that the mean writhes of the Boltzmann populations of polymers forming different composite knots are practically equal to the writhes of the ideal configurations of these knots. We have shown¹ that the mean writhe of a Boltzmann population of polymers forming a given prime knot is equal to the writhe of the ideal configuration of the knot; taking this into account, the data in Table 1 show that the additivity of writhe applies also to the time-averaged writhe values of thermally distorted configurations of the factor knots making up a given composite knot. Of course, in individual configurations of thermally distorted polymeric chains, the writhe of the composite knots varies. Random thermal distortions, however, are equally likely (for reasons of symmetry) to introduce regions of left- or right-handed winding in the knots. Thus, on average, the writhes of random knots, either created by thermal agitation or generated on a cubic

Table 1 Writhe of ideal and random configurations of composite knots

Composite knot type	Writhe of ideal form	Mean writhe of the population of knotted DNA molecules in solution
$+3_1\#3_1$	6.82	6.78
$+3_1\#-3_1$	0.00	0.00
$4_1\#3_1$	3.39	3.42
$+5_1\#3_1$	9.67	9.61
$+5_1\#-3_1$	2.84	3.00
$+5_2\#3_1$	7.98	7.96
$+5_2\#-3_1$	1.17	1.14
$4_1\#4_1$	0.05	0.03
$+6_1\#3_1$	4.55	4.40
$+6_1\#-3_1$	-2.25	-2.15
$+6_2\#3_1$	6.15	6.19
$+6_2\#-3_1$	-0.55	-0.54

We note the good correspondence between the mean writhes of the simulated Boltzmann populations of randomly distorted, nicked DNA molecules forming different composite knots and the writhe values of the corresponding ideal configurations of the composite knots. Simulations of equilibrium configurations of composite and prime knots were performed for DNA molecules (5,400 base pairs long) in low salt solution. Taking into account statistical errors in sampling the configurational space of randomly distorted knots and the error in calculating the writhe with the Gauss integral⁷, the data in the second and third columns can be regarded as equal. See ref. 8 for further details on the simulation procedure.

lattice⁶, correspond to the writhe of the ideal representation of a given knot.

The principal characteristic of an ideal knot is the ratio of its length to diameter, which we call the relative length. This ratio is a good measure of the complexity of a knot; the higher the relative length is, the more complicated the knot is. Although the relative lengths of the constituent ideal factor knots do not sum to that of their ideal composite knot, the total exceeds the relative length of the composite knot by roughly the same value regardless of the type of composite knot; see Fig. 2b. Closer inspection of Fig. 1 reveals the reason for this subadditivity. If we assume that the diameter of the tube forming the composite knots shown in the figure is 1 cm, it should not be too difficult to see that in the formed composite knots each factor knot loses one half of a quasi-circular arc of 1 cm radius and gains one straight segment of ~ 1 cm length. The arcs which are eliminated by the fusion are, in fact, somewhat longer than π , as the winding angles are not equal to 90° . Thus, in total, the length decrease is usually somewhat greater than the diameter multiplied by $2\pi - 2$. Because the junction points of factor knots can have different local structures for various types of composite knots (see Fig. 1), the lengths of eliminated arcs and new linking segments may vary. This produces small deviations from subadditivity with a constant difference.

As already mentioned, the loops of symmetric factor knots such as 3_1 or 4_1 are identical. Thus, irrespective of which loop is used to fuse two such knots, no major reshuffling of the knot is required to reach the ideal configurations of $3_1\#3_1$, $3_1\#4_1$ or $4_1\#4_1$ composite knots (see Fig. 1 for details of this nomenclature). However, there are several ways of arranging composite knots constructed from multiple repetitions of 3_1 or 4_1 factor knots, and a major reshuffling of the whole length of the composite knot is needed to convert from one form into another. Consider the composite knot formed from four 3_1 knots. The factor knots can be joined circularly (Fig. 1, bottom row), linearly (Fig. 3a) or in a star-like configuration with the loops of a central trefoil linked to the three other trefoils (Fig. 3b). To change from the linear to the star-shaped configuration of a composite knot, energy is required to loosen the knot. By providing more length or by shrinking the diameter, the factor knots can be rearranged in such a way that a different isomer of the composite knot will appear. Composite knots made up of multiple repeats of four-noded knots recall hydrocarbon molecules, as the ideal form of

the four-noded knot has roughly tetrahedral symmetry, and it is possible to create star-shaped arrangements where the central four-noded knot coordinates to four other knots. □

Received 19 March; accepted 13 May 1997.

1. Katritch, V. *et al.* Geometry and physics of knots. *Nature* **384**, 142–145 (1996).
2. Stasiak, A., Katritch, V., Bednar, J., Michoud, D. & Dubochet, J. Electrophoretic mobility of DNA knots. *Nature* **384**, 122 (1996).
3. Rolfsen, D. *Knots and Links* (Publish or Perish, Berkeley, CA, 1976).
4. Fuller, F. B. The writhing number of a space curve. *Proc. Natl Acad. Sci. USA* **68**, 815–819 (1971).
5. Fuller, F. B. Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc. Natl Acad. Sci. USA* **75**, 3557–3561 (1978).
6. Janse van Rensburg, E. J., Orlandini, E., Summers, D. W., Tesi, M. C. & Whittington, S. G. The writhe of knots in the cubic lattice. *J. Knot Theory Ramif.* **6**, 31–44 (1997).
7. Calugareanu, G. L'intégral de Gauss et l'analyse des noeuds tridimensionnels. *Rev. Math. Pur. Appl.* **4**, 5–20 (1959).
8. Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank-Kamenetskii, M. & Cozzarelli, N. R. Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.* **227**, 1224–1243 (1992).
9. Pieranski, P. Search of ideal knots. *ProDialog* **5**, 111–120 (1996).

Acknowledgements. This work was supported by the US Public Health Service, the Polish Committee of Scientific Research, the Swiss National Foundation and by Foundation Herbet, University of Lausanne. We thank R. Scharein for making available his KnotPlot program.

Correspondence should be addressed to A.S. (e-mail: Andrzej.Stasiak@lau.unil.ch).

The dynamics of partially extended single molecules of DNA

Stephen R. Quake*, Hazen Babcock & Steven Chu

Physics Department, Stanford University, Stanford, California 94305, USA

The behaviour of an isolated polymer floating in a solvent forms the basis of our understanding of polymer dynamics^{1,2}. Classical theories describe the motion of a polymer with linear equations of motion, which yield a set of 'normal modes', analogous to the fundamental frequency and the harmonics of a vibrating violin string. But hydrodynamic interactions make polymer dynamics inherently nonlinear, and the linearizing approximations required for the normal-mode picture have therefore been questioned¹. Here we test the normal-mode theory by measuring the fluctuations of single molecules of DNA held in a partially extended state with optical tweezers. We find that the motion of the DNA can be described by linearly independent normal modes, and we have experimentally determined the eigenstates of the system. Furthermore, we show that the spectrum of relaxation times obeys a power law.

The starting point for describing the motion of a polymer is the Rouse model of beads interconnected with gaussian springs. If the interactions between different segments (beads) of the polymer R_n and R_m (n and m denote the order of the beads along the length of the polymer) are localized to nearest neighbours, each of the segments R_n satisfies a linear differential equation decoupled from the other segments. This linear system allows one to describe the dynamics in terms of a set of normal modes with mode amplitudes X_p defined as $X_p = (1/N) \int_0^N R_n \cos(\pi p n / N) dn$ (where p is an integer denoting the p th mode and N is the total number of beads). Associated with the normal modes is a set of relaxation times, τ_p , which describe the decay in the time correlations of the normal-mode amplitudes $\langle X_p(t) X_q(0) \rangle \sim \delta_{pq} e^{-t/\tau_p}$ (where q is the q th mode).

Hydrodynamic interactions between segments of the polymer were first included by Zimm³ by assuming the surrounding fluid is incompressible and obeys the Navier–Stokes equation (known as the Kirkwood approximation⁴). The resulting coupling of each polymer segment R_n to all of the other segments R_m leads to a

nonlinear set of differential equations as the coupling depends on the instantaneous configuration of the polymer. These equations can be linearized by first averaging the distance $|R_m - R_n|$ between segments over the distribution of accessible configurations. This so-called 'pre-averaging' approximation linearizes the set of coupled differential equations and allows one to construct a set of normal modes similar to the Rouse modes.

It has been argued that a normal-mode description of a polymer can be only a rough approximation of the true dynamics and may not have any fundamental validity¹. Apart from questioning the soundness of the pre-averaging approximation, other nonlinear effects such as excluded-volume effects, forbidden crossings, and knots were not included in Zimm's treatment. These and other nonlinear effects have led de Gennes¹ and others to emphasize dynamical scaling laws rather than a normal-mode description. Scaling laws link relaxation rates to the spatial scale of interest, but do not require knowledge of the precise shape of the relaxation spectrum. On the other hand, the normal-mode concept requires that the relaxation spectrum be described by a set of discrete frequencies.

The analysis of the relaxation of polymers perturbed from equilibrium can yield a frequency spectrum by taking the inverse Laplace transform of the relaxation spectrum, but these methods do not provide an unambiguous decomposition into discrete frequencies^{5,6}. Dynamic light-scattering experiments^{7,8}, viscoelastic and oscillatory flow birefringence studies^{9–12} and transient electric birefringence experiments^{13,14} probe relaxation rates at size scales considerably smaller than the overall coil size, but these measurements represent an average over what could be a continuous spectrum of relaxation rates. Dielectric relaxation of dipole-inverted *cis*-polyisoprene has been used to measure the first two normal-mode eigenfunctions, and a tentative extrapolation was used to derive the third^{15–17}. However, this method does not measure a relaxation mode spectrum, and does not discriminate between linear and nonlinear effects.

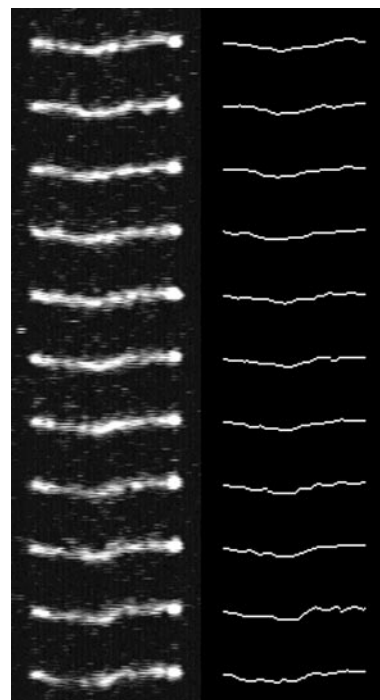


Figure 1 Video images of single molecules of DNA held at an extension of 12 μm . The column on the left shows a series of successive raw video images; the column on the right shows the corresponding parametrizations used in the data analysis.

* Present address: Department of Applied Physics, Caltech, MS 128-95, Pasadena, California 91125, USA.

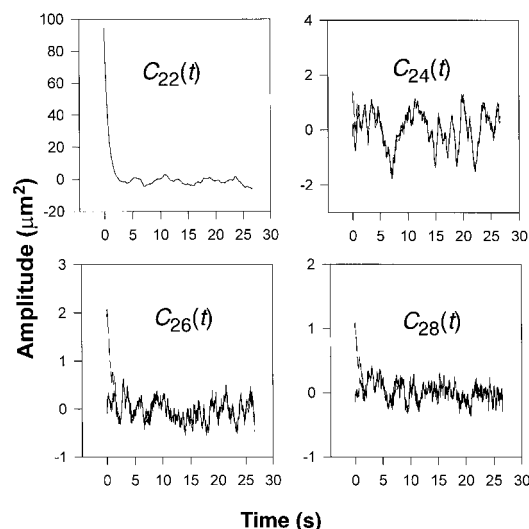


Figure 2 Comparison of the sine basis and the normal-mode basis for several matrix elements of DNA held at 10 μm extension. The correlation function $C_{pq}(t)$ represents eigenbasis correlation matrix $\langle X_p^{(y)}(t)X_q^{(y)}(0) \rangle$ for solid lines and the sine basis correlation matrix $\langle U_p^{(y)}(t)U_q^{(y)}(0) \rangle$ for dashed lines. We note that in the normal-mode basis, there are no off-diagonal elements above the level of noise fluctuations. The increasing frequency of the characteristic fluctuations for the cross-correlation of mode 2 with successively higher modes may be seen in this figure.

We have tested the supposition that polymer motion can be described by a set of normal modes by recording the images of partially extended 20- μm -long pieces of DNA driven by brownian motion. The DNA is suspended in solution by attaching 0.3- μm polystyrene spheres to the ends of the molecule that serve as 'handles' for two optical tweezers¹⁸. The DNA, fluorescently stained with the dye YOYO-1 (Molecular Probes, Eugene, OR) and illuminated by 488-nm light, was held in a partially extended state with optical tweezers while the images were recorded by video microscopy (Fig. 1). Analysis of the static properties of partially extended polymers (to be presented elsewhere) shows that intercalation by YOYO-1 increases the length of λ DNA from 16.4 to 20 μm , and increases the persistence length from 50 to 66 nm.

The optical images were converted into parametrized descriptions of the DNA location as described below. If the line joining the two beads defines the x -axis, we analyse the polymer configuration in terms of displacements $R_n^{(y)}(t)$ along the y -axis. The boundary conditions of fixed end points suggests that a good first approximation to the projection of the normal modes describing the y -motion would be the sine functions with amplitudes $U_p^{(y)} = (1/N)\int_0^N R_n^{(y)}(t) \sin(\pi p n/N) dn$. Normal modes have also been calculated for boundary conditions where only one end of the polymer is fixed¹⁹.

After computing $U_p^{(y)}(t)$ by taking a sine transform of $R_n^{(y)}(t)$ (ref. 20), we calculated the cross-correlation matrix $\langle U_p^{(y)}(t)U_q^{(y)}(0) \rangle$ for the first eight sine modes. This matrix is not diagonal and has off-diagonal relaxations (Fig. 2). For example, for 10 μm extension the amplitude of the $\langle U_2^{(y)}(0)U_6^{(y)}(0) \rangle$ matrix element is 1.9 μm^2 , as compared to the 86.2 μm^2 for the $\langle U_2^{(y)}(0)U_2^{(y)}(0) \rangle$ element and 16.0 μm^2 for the $\langle U_6^{(y)}(0)U_6^{(y)}(0) \rangle$ element. Random noise in the measurement can be estimated from

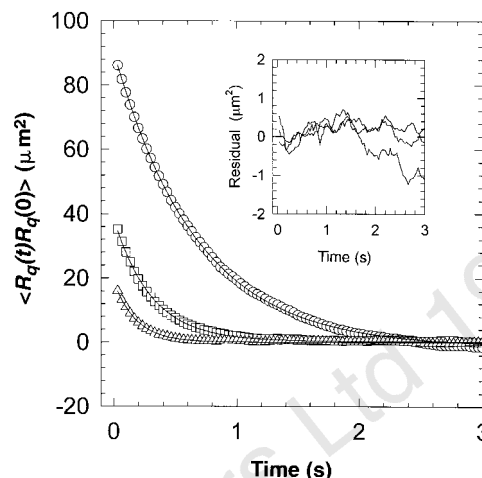


Figure 3 Relaxation curves and exponential fits for $q = 2, 3$, and 4 (circles, squares and triangles, respectively) at extension 10 μm . Inset, residuals from the fits, showing a small systematic deviation from pure exponential relaxation. There is a measurement error in assigning the parametrization of the polymer; this causes a systematic error in the $t = 0$ correlation matrix. If $R_n(t)$ denotes the actual position of the polymer and $\epsilon_n(t)$ denotes the measurement error, then the measured quantity is $R_n(t) + \epsilon_n(t)$. $\langle (R_q(t) + \epsilon_q(t))(R_q(0) + \epsilon_q(0)) \rangle = \langle R_q(t)R_q(0) \rangle + \langle \epsilon_q(t)\epsilon_q(0) \rangle$, where the cross terms $\langle R_q(t)\epsilon_q(0) \rangle + \langle \epsilon_q(t)R_q(0) \rangle = 2\langle R_q \rangle \langle \epsilon_q \rangle$ are zero. The measurement error should be independent of time, so $\langle \epsilon_q(t)\epsilon_q(0) \rangle$ vanishes for $t \neq 0$. We checked this by fitting the relaxation function to two exponentials; one relaxation time was always degenerately small, while the amplitudes were in agreement with our other estimations of the signal and noise. Thus, the intrinsic noise has a systematic effect on only the $t = 0$ point of the correlation function. That point was discarded before fitting the relaxation to an exponential function.

the r.m.s. fluctuations of the matrix elements after all correlations have disappeared. The r.m.s. fluctuations of $\langle U_2^{(y)}(t)U_2^{(y)}(0) \rangle$, $\langle U_2^{(y)}(t)U_6^{(y)}(0) \rangle$ and $\langle U_6^{(y)}(t)U_6^{(y)}(0) \rangle$ were computed between $6.7 < t < 26.7$ seconds, yielding 0.22 μm^2 , 2.05 μm^2 and 0.07 μm^2 , respectively. The off-diagonal relaxations are therefore statistically significant and show that the sine functions are not the normal modes of the system.

Zimm recognized that the sine basis is only an approximation to the exact eigenbasis, and calculated higher-order corrections²¹. Similarly, we sought an improved eigenbasis constructed for the sine basis $u_k(x, t) = U_k^{(y)}(t) \sin(k\pi x/L)$, where L is the distance between the end points of DNA, in the form $x_i^{(y)}(x, t) = \sum_k B_{ik} x_k^{(y)}(x, t)$. The condition for the time correlations of the normal-mode amplitudes $X_i(t)$ becomes

$$\begin{aligned} \langle X_i(t)X_j(0) \rangle &= \left\langle \left(\sum_k B_{ik} U_k(t) \right) \left(\sum_l B_{jl} U_l(0) \right) \right\rangle \\ &= \sum_{k,l} B_{ik} B_{jl} \langle U_k(t)U_l(0) \rangle \\ &= \delta_{ij} A_i e^{-t/\tau_i} \end{aligned}$$

We solved for the matrix B by diagonalizing the cross-correlation matrix for $t = 1/30$ s (one video frame). Figure 2 shows that the resulting basis, given in Table 1, satisfies the orthogonality condition. Time cross-correlations of different mode amplitudes were indistinguishable from noise. If normal modes did not exist, the transformation matrix would be a function of time and we would expect to see non-zero correlation functions in a time t between 1/30 s and the mean of τ_p and τ_q . In addition to demonstrating the

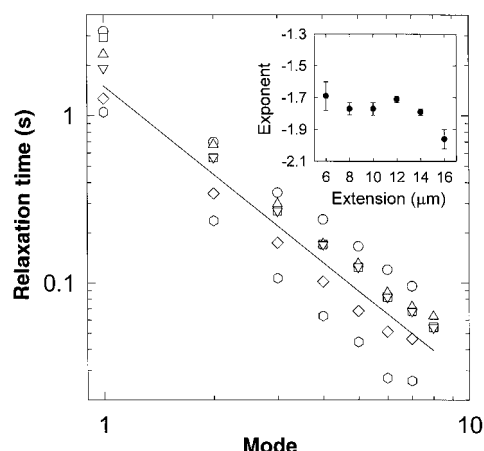


Figure 4 Relaxation time versus mode (q) from fitting to single exponentials (circle, 6 μm extension; square, 8 μm ; down triangle, 10 μm ; up triangle, 12 μm ; diamond, 14 μm ; hexagon, 16 μm). The line has a slope of -1.7 . There is a dependence of relaxation time on extension: the greater the extension, the higher the 'string tension' and the faster the relaxation. The ultimate limit of how many modes we could resolve was determined by both the temporal and spatial resolution of the system. The time resolution of one video frame ($=0.033$ s) limited the most extended polymers to eight modes. The spatial resolution was lowest for the least extended polymers, as the 'slack' in the DNA increases the width of the image. For the smallest extension (6 μm) the average full-width at half-maximum of the image in the transverse direction was 0.71 μm , indicating a cut-off after eight modes. Inset, fitted exponents for the mode structure as a function of extension.

orthogonality of the normal modes, we show in Fig. 3 that the diagonal relaxation functions $\langle X_p^{(y)}(t)X_p^{(y)}(0) \rangle$ are well described by a single exponential decay as required for a linear dynamical system.

Figure 4 shows that the relaxation times τ_p scale as a power law over almost two decades of time. The Zimm model for a coiled polymer predicts an exponent of -1.5 for a Θ -solvent² and -1.8 for a good solvent. DeGennes has predicted²² that a highly extended polymer will have an exponent of -2.0 . Caution must be used in comparing the observed power-law scaling to dynamical models that were derived for conditions where the end points are not constrained. Although it is known that our buffer conditions correspond to a good solvent for DNA²³, it has been suggested that an extended polymer will behave as in an ideal solvent on moderate to long length scales such as we measured^{24,25}. Also, the pre-averaged distances for a partially extended polymer are different than for a free polymer. For an unstretched chain, $\langle |R_n - R_m| \rangle \sim |n - m|^{0.5}$ for an ideal solvent and $|n - m|^{0.6}$ for a good solvent leading to relaxation times that scale as $\tau_p \approx p^{-1.5}$ or $p^{-1.8}$, respectively. For partially extended polymers, at distances $|n - m| \gg 1$, we expect $\langle |R_n - R_m| \rangle_{\text{pre-averaged}}^{-1} \approx |n - m|^{-1.0}$, but for $n \approx m$, $|n - m|^{-\alpha}$ where the exponent α might be some value between 0.5 and 1.0. Clearly, a calculation of the pre-averaged hydrodynamic tensor for our geometry will have to be done before one compares these results to a particular dynamical model. $\square$

Methods

Sample preparation. λ -Phage DNA was attached to streptavidin-coated beads via biotinylated oligonucleotides that hybridized to the single-stranded ends of λ DNA. The λ DNA and one of the oligonucleotides were mixed in equal molar amounts in 50 mM NaCl-TE (10 mM Tris-Cl, 1 mM EDTA, pH 8.0), heated to 70 $^{\circ}\text{C}$ for 10 min, and allowed to anneal to room temperature for 2–3 h. The

Table 1 Normal-mode eigenbasis

1.000	0.000	-0.011	-0.011	-0.019	-0.004	-0.008	0.002
0.000	0.999	0.029	-0.018	-0.006	-0.024	-0.005	-0.011
0.010	-0.027	0.997	0.023	-0.043	0.002	-0.044	-0.007
0.012	0.020	-0.024	0.997	0.020	-0.025	-0.021	-0.046
0.019	0.007	0.038	-0.044	0.996	0.010	-0.073	0.000
0.005	0.024	-0.004	0.027	-0.017	0.999	0.003	-0.041
0.010	0.003	0.047	0.023	0.074	-0.001	0.995	-0.027
0.000	0.012	0.009	0.045	-0.007	0.040	0.049	0.998

Columns show the normal modes expressed in terms of a sum of sine vectors for the 10- μm extension. The rows are the mode numbers. For example, column 8 shows that the eighth normal mode has an amplitude 0.998 of $\sin(8\pi n/N)$, -0.027 of $\sin(7\pi n/N)$ and -0.041 of $\sin(6\pi n/N)$.

opposite oligonucleotide was mixed with streptavidin beads in about 2 $\times$ excess to streptavidin binding sites, and allowed to incubate for 8 h before being washed three times by centrifugation. Equal molar amounts of biotinylated DNA, oligo-coated beads, and streptavidin beads were mixed and allowed to incubate for 12 h at room temperature in 50 mM NaCl-TE. The resulting bead-DNA-bead complexes were diluted by a factor of 10 into a staining solution contain TE, 2 mM NaCl, 100 nM YOYO, 1% β -mercaptoethanol, and 0.1% Tween-20. After 45–60 min of staining, the complexes were diluted by another factor of 10 into a high-viscosity TE buffer containing a final concentration of 71% (w/v) glycerol, 2 mM NaCl, and 0.1% Tween-20. To retard photobleaching of the dye, 0.1 mg ml⁻¹ glucose oxidase, 0.04 mg ml⁻¹ catalase, and 0.4% glucose were added.

Experimental procedure. The optical tweezers were made by focusing two independently controllable infrared (1.064- μm) Nd:YAG laser beams through a $\times 63$, 1.4 numerical aperture microscope in a home-made microscope with a measured resolution of 0.25 μm . The DNA was held 10 μm below the surface of a 30- μm -deep cell made from a microscope slide and coverslip sealed with epoxy. The fluorescent images were recorded with an image-intensified CCD camera with a video rate of 30 Hz. Because the fundamental relaxation time of a polymer is proportional to viscosity, we increased the effective time resolution of the experiment by increasing the viscosity from 0.01 to 0.2 poise with the addition of 71% glycerol (w/v). We took data from 37 different molecules held at six different extensions ranging between 30% and 80% of the full length of the molecule. An average of 24 min (43,200 video frames) of data were taken at each extension.

Data analysis. Spatial noise in the images were filtered out by convolving each image with a gaussian of width 0.42 μm before finding the fluorescence maxima. Using continuity of the DNA, an algorithm tracked the fluorescent peak from one bead to the other and back again. As the DNA fluctuated in and out of the focal plane, parts of its image broadened and occasionally disappeared. The fitting program discarded those points; 98% of the data was retained, and a visual and manual check ensured that the fitting program was not giving skewed values. Although details of the DNA configuration below the 0.25 μm resolution of the microscope are not seen, we estimate that the centre of the image is determined with an uncertainty of ± 1 pixel which is ± 0.14 μm . We showed that the resulting 'pixelation' of the images does not affect the data analysis by artificially increasing the linear pixel size by a factor of two; the results of the subsequent data analysis did not change. The beads are free to rotate in the traps; this allows the ends of the polymer a small amount of motion not greater than the radius of the bead. Also, the beads absorb some of the YOYO and fluoresce brightly; this causes blooming on the video image and prevents us from locating the precise end points of the DNA. As we infer the locations of those end points from the locations of the beads, there is a small systematic overestimation of the distance of the DNA from its equilibrium position as it fluctuates above and below the imaginary line which connects the two beads. We put that extra degree of freedom of the end points into a computer simulation of a bead-spring model with the exact hydrodynamic tensor. For exponential fits, the power-law exponent of the simulated mode structure with mobile end points was 2.5% smaller than the value for fixed end points, with a 2.5% statistical error on the fit. Thus the rotation of the beads has only a small systematic effect on the data. One can correct for this effect by multiplying the measured exponents by 1.025.

Received 12 September 1996; accepted 13 May 1997.

1. de Gennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY, 1979).
2. Doi, M. & Edwards, S. *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
3. Zimm, B. H. Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.* **24**, 269–278 (1956).
4. Kirkwood, J. & Riseman, J. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *J. Chem. Phys.* **16**, 565–573 (1948).
5. Perkins, T., Quake, S., Smith, D. & Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* **264**, 822–826 (1994).
6. Pecora, R. DNA—A model-compound for solution studies of macromolecules. *Science* **251**, 893–898 (1991).
7. Sorlie, S. & Pecora, R. A dynamic light-scattering study of 4 DNA restriction fragments. *Macromolecules* **23**, 487–497 (1990).
8. Adam, M. & Delsanti, M. Dynamical properties of polymer solutions in good solvent by Rayleigh scattering experiments. *Macromolecules* **10**, 1229–1237 (1977).
9. Sahooani, H. & Lodge, T. P. Onset of excluded-volume effects in chain dynamics. *Macromolecules* **25**, 5632–5642 (1992).
10. Amelar, S. *et al.* Dynamic properties of low and moderate molecular-weight polystyrenes at infinite dilution. *Macromolecules* **24**, 3505–3516 (1991).
11. Sammler, R. L. *et al.* Polydispersity effects on dilute-solution dynamic properties of linear homopolymers. *Macromolecules* **23**, 2388–2396 (1990).
12. Johnson, R., Schrag, J. & Ferry, J. Infinite-dilution viscoelastic properties of polystyrene in θ -solvents and good solvents. *Polym. J.* **1**, 742–749 (1970).
13. Hong, M. K. *et al.* Internal dynamics of DNA probed by transient electric birefringence. *Phys. Rev. Lett.* **68**, 1430–1433 (1992).
14. Lewis, R., Pecora, R. & Eden, D. Transient electric birefringence measurements of the rotational and internal bending modes in monodisperse DNA fragments. *Macromolecules* **19**, 134–139 (1986).
15. Watanabe, H., Yao, M. & Osaki, K. Comparison of dielectric and viscoelastic relaxation behavior of polyisoprene solutions—coherence in subchain motion. *Macromolecules* **29**, 97–103 (1996).
16. Watanabe, H., Yamada, H. & Urukawa, O. Dielectric-relaxation of dipole-inverted cis-polyisoprene solutions. *Macromolecules* **28**, 6443–6453 (1995).
17. Watanabe, H., Urukawa, O. & Kotaka, T. Slow dielectric-relaxation of entangled linear cis-polyisoprenes with asymmetrically inverted dipoles. 1. Bulk systems. *Macromolecules* **26**, 5073–5083 (1993).
18. Ashkin, A., Dziedzic, J., Bjorkholm, J. & Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.* **11**, 288–290 (1986).
19. Marciano, Y. & Brochard-Wyart, F. Normal-modes of stretched polymer-chains. *Macromolecules* **28**, 985–990 (1995).
20. Press, W., Teukolsky, S., Vetterling, W. & Flannery, P. *Numerical Recipes* (Cambridge Univ. Press, New York, 1994).
21. Zimm, B., Roe, G. & Epstein, L. Solution of a characteristic value problem for the theory of chain molecules. *J. Chem. Phys.* **24**, 279–280 (1956).
22. De Gennes, P. G. Coil-stretch transition of dilute flexible polymers under ultrahigh velocity gradients. *J. Chem. Phys.* **60**, 5030–5042 (1974).
23. Smith, D., Perkins, T. & Chu, S. Dynamical scaling of DNA diffusion-coefficients. *Macromolecules* **29**, 1372–1373 (1996).
24. Pincus, P. Excluded volume effects and stretched polymer chains. *Macromolecules* **9**, 386–388 (1976).
25. Pincus, P. Dynamics of stretched polymer chains. *Macromolecules* **10**, 210–213 (1977).

Acknowledgements. We thank T. T. Perkins and D. E. Smith for contributions and discussions in the early stages of this work, and B. Zimm, R. Larson, T. Lodge, G. Fuller, E. Shaqfeh, S. Doniach and W. Volkmuth for discussions. This work was supported in part by the Air Force Office of Scientific Research, the National Science Foundation, and the Human Frontier Science Program. S.R.Q. was supported in part by an NSF fellowship.

Correspondence and requests for materials should be addressed to S.C. (e-mail: schu@leland.stanford.edu).

Multiple episodes of aridity in southern Africa since the last interglacial period

Stephen Stokes^{*†}, David S. G. Thomas[‡] & Richard Washington^{*}

^{*} School of Geography, University of Oxford, Mansfield Road, Oxford OX1 3QJ, UK

[†] Research Laboratory for Archaeology and the History of Art, University of Oxford, Mansfield Road, Oxford OX1 3TB, UK

[‡] Sheffield Centre for International Drylands Research, Department of Geography, University of Sheffield, Sheffield S10 2TN, UK

There is generally a dearth of evidence of the nature of Quaternary climate change within desert systems, which has limited previous interpretations of past environmental change at low latitudes. The Last Glacial Maximum has previously been identified as the peak of Late Quaternary aridity, when desert systems expanded to five times their present extent^{1–3}, and low-latitude aridity has been described for previous glaciations⁴. But little evidence has been derived directly for large desert basins, particularly southern

Africa. Here we report new chronological (optical dating) evidence of arid episodes recorded in aeolian sediments from the Mega Kalahari sand sea. Episodic aeolian activity is recorded at the northeastern desert margin, whereas more sustained activity is evident from the southwestern desert core. Several significant arid events are apparent since the last interglacial period, with dune-building (arid) phases at ~95–115, 41–46, 20–26 and 9–16 kyr before present. Existing atmospheric general circulation model simulations and independent palaeoclimate data indicate that the changes in aridity are related to changes in the northeast–southwest summer rainfall gradient, which are in turn related to sea surface temperatures in the southeastern Atlantic Ocean.

The 2.5×10^6 km² Mega Kalahari is the world's most continuous area of aeolian sand sea which is dominated by variably degraded and pedogenically modified, currently inactive, linear dunes⁵ (Fig. 1). The development and evolution of the extensive Kalahari linear dune systems has not until now been directly dated. Phases of aridity and dune development have previously been inferred from gaps within subcontinental humid chronologies^{6,7}, or by assumed mirroring of Northern Hemisphere conditions⁸. We have determined the timing of dune building and associated palaeoaridity along the present day northeast–southwest summer rainfall gradient by optically dating⁹ dune sediments from geographically and morphologically wide-ranging aeolian deposits. Past changes in aeolian dynamism is assumed to be related to shifting rainfall intensities along the gradient (Fig. 1).

The effect of this northeast–southwest rainfall gradient on aeolian activity provides an opportunity to determine the sensitivity of the climate system to changing boundary conditions since the last interglacial. The climatology of this gradient is an infrequently studied element of the general circulation of the subcontinent, yet one that is accurately captured in all present-day leading general circulation models^{10,11}. It originates from the Earth's largest cross-continental zonal asymmetry of tropical convection; the inter-tropical convergence zone (ITCZ) in the southwestern Indian Ocean occupies the most southerly location of any ocean at 23° S, whereas tropical convection in the eastern Atlantic Ocean is seldom found south of 5° N (ref. 11). Convection over southern Africa aligns almost meridionally over the subcontinent. Differences in the latitude of convection in the adjacent oceans relates directly to sea surface temperatures (SSTs).

The southwestern Indian Ocean is the warmest ocean at 23° S whereas the southeastern Atlantic is the coldest (present-day January mean of 27.5 and 21.6 °C, respectively). The extreme SST contrast across the subcontinent determines the southern African rainfall gradient (Fig. 1). Wetter conditions east of the Kalahari result from disturbances in the tropical easterlies associated with the Mascarene anticyclone. Westward propagation of easterly waves off the African subcontinent is blocked by cold stable air overlying the Benguela Current. Conditions west of the wave axis are therefore dry. Intense subsidence into subtropical anticyclones dominate the climatology of the austral winter, ensuring aridity. The elevated southern African land mass (2,000 m) cools through longwave emission, thereby providing an interhemispheric sink for Asian monsoon outflow. Few deserts provide a suite of controls on the rainfall regime as diverse as the Kalahari.

Although the controls on aeolian sedimentation, particularly in the case of linear dune development¹², may be complex we believe that our chronology of dune construction reflects periods of regionally enhanced aridity. In the northeastern Kalahari, mean rainfall today exceeds 400 mm yr⁻¹, and the pedogenically modified and vegetated surfaces of degraded linear dunes prohibits aeolian activity. In the southwestern Kalahari the present mean annual rainfall is 150–200 mm. Episodic, localized aeolian activity is confined to dune crests and controlled by variations in vegetation cover induced by interannual variations in rainfall and intensive grazing, but is in any case significantly restricted by present-day low

wind energy conditions¹³. Conditions throughout the Kalahari today, therefore, are neither arid nor windy enough to permit wholesale dune construction or reworking; the southwestern region is, however, somewhat closer to a threshold of dune reactivation than the northeastern region.

Samples for optical dating⁹ (Table 1) were collected from linear dunes and ancillary aeolian bedforms (mainly lunette dunes and dune patches) from the two major dunefields in the sand sea (Fig. 1). Where feasible, a sequence of vertically spaced samples was analysed, facilitating assessments of the rate and continuity of aeolian accumulation. The four types of aeolian deposit sampled in the southwestern Kalahari yielded contrasting age assessments spanning the past in 30 kyr (Fig. 2; Table 1). Samples from basal sediments within linear dunes consistently yielded ages exceeding 20 kyr, whereas analyses on sands from the body of dunes describing the main linear forms of the area generally indicate depositional ages in the range 10–20 kyr. Samples 945/1 & 948/1 from the basal sands at the contact of non-aeolian strata demonstrate that aeolian activity from 22 to 28 kyr ago operated to the full depth of the sand sea, precluding the preservation of older dunes. The maximum extent of the Kalahari to the southwest is constrained by three optical dates for a dune which invaded the Orange River valley at 13.2 ± 1.5 kyr ago. Based on palaeoecological evidence^{7,14}, we infer either reduction or cessation of dune activity in the periods 22–16 and 6–10 kyr ago.

Dune construction during the Holocene epoch appears to have been less intensive, resulting only in localized dune deposits. Mid-Holocene (~6 kyr ago) aridity is demonstrated from superficial reticulate dunes (site 946; Table 1), and Late Holocene aridity (≤ 2 kyr ago) from lunette deposition (site 949), and localized disturbance patches (site 947). The timing of the three post-LGM periods of aridity accord well with gaps in the northern Cape humid chronologies¹⁵.

Aeolian geomorphic units sampled within the northeastern Kalahari comprised either weathered and degraded linear dunes or poorly defined dune mounts (Table 1). In contrast to the southwest, linear dunes in the northeast are larger and substantially older features, and there is little evidence for Holocene deposition. Detailed analysis of a 6-m-deep linear dune profile delineated major phases of dune construction at ~95–115, 41–46 and 20–26 kyr ago (Fig. 2c). The sequence is suggestive of relatively short-lived (~5–20 kyr) periods of non-deposition, landscape stability and increased net moisture which are recorded as subtle changes in pedogenesis and organic-matter content. Limited late glacial and Early Holocene aeolian activity (10–16 kyr ago) is also identified in this area, typically superimposed on the principal linear dune pattern. Independent palaeoenvironmental evidence^{16,17} supporting increases in moisture during the period 20–15 kyr ago implies that this latest event is separate from the 20–26 kyr event. Middle and later Holocene dune reactivation is absent from the area.

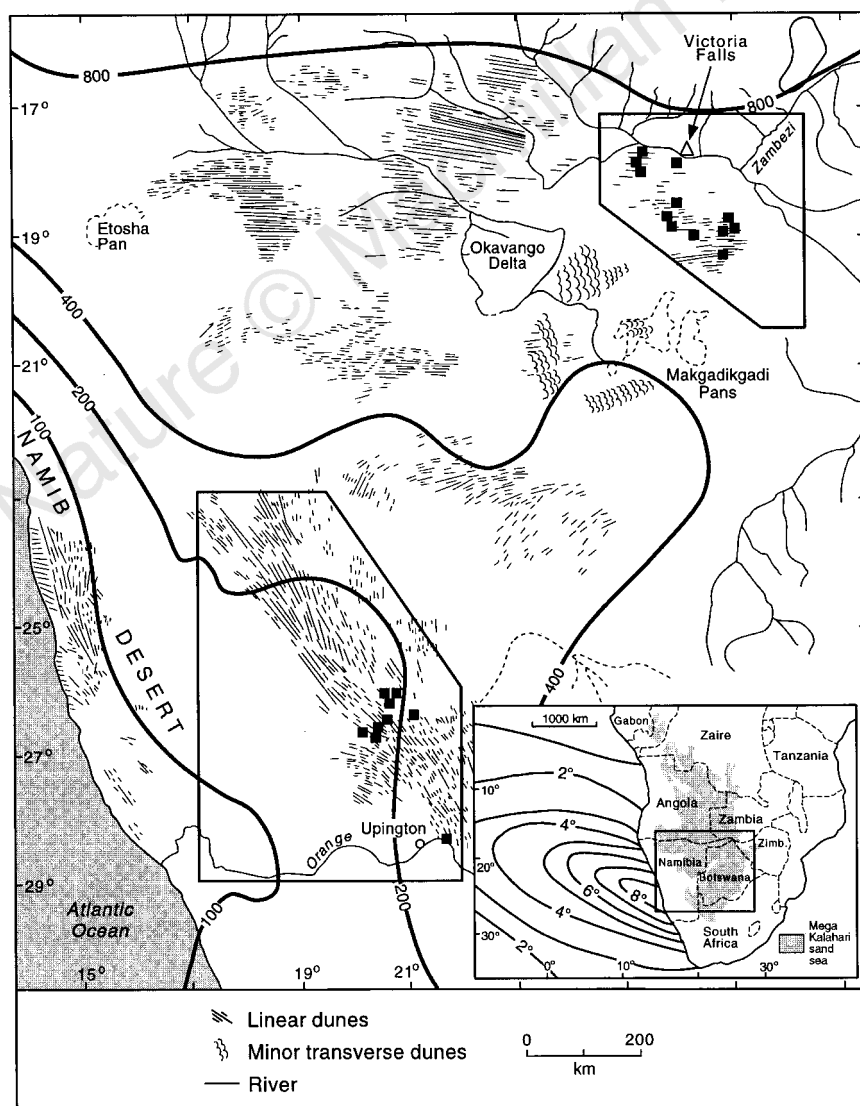


Figure 1 Sample sites within the linear dune systems are shown by filled squares in the two boxed areas (which are shown in detail in Fig. 2). Isohyets (contours) show present-day mean annual rainfall in the region and demonstrate the northeast-southwest regional rainfall gradient in the continental interior. Intense present-day aridity and true desert conditions are restricted to the geographically isolated coastal Namib Desert. Inset, location of the study area in the context of the 2.5 × 10⁶ km² Mega Kalahari sand sea and symmetrical Indian/Atlantic Ocean mean (1951–80) temperature differences (°C) for the January period. The temperature-difference contours show the cold anomaly of the Atlantic Ocean relative to the Indian Ocean. Cold Atlantic water is maintained through 'symbiosis' between ocean and atmosphere; the subtropical oceanic gyre is driven through wind stress by the South Atlantic anticyclone which benefits, in turn, from the cold ocean surface it helps to maintain.

In both the northeastern and southwestern Kalahari there is evidence of extensive linear dune field construction (persistent aridity) during the later Quaternary, which was succeeded in the southwest by episodes of minor, spatially confined, dune emplacement. The record of linear dune development in the northeast, where major dune construction periods extend to the Pleistocene/Holocene boundary, is of considerably greater antiquity than that preserved in the southwest.

A preponderance of studies of North African Quaternary palaeoclimates based on terrestrial and oceanic evidence have failed to identify a single forcing mechanism of climate changes. Instead, the preserved records support influences of both precessional variations at the 19–23 kyr periodicities enhancing summer monsoonal rainfall regimes—particularly before the onset of high-latitude Quaternary glacial conditions¹⁸—and the 41- and 100-kyr obliquity and

eccentricity periodicities as influencing continental aridity, modulated at least in part by related changes (reductions) in SSTs^{4,19,20}. Given the paucity of high-resolution marine records for the oceans adjacent to Southern Africa, theories of large-scale forcing of southern African climates have focused on generally more recent (≈ 25 kyr), sparsely distributed, discontinuous terrestrial records and some coastal sequences^{21,22}.

Changes in the moisture regime which have influenced dune reactivations relate to changes in the extent of summer precipitation, which we interpret as changes in convection along the north-east–southwest gradient. These, in turn, relate to the controls on this gradient, such as SSTs, identified earlier. To determine the sensitivity of this gradient to SSTs, we pursued a set of GCM experiments where the United Kingdom Meteorological Office (UKMO) atmospheric GCM HADAM2A (resolution 2.5° latitude,

Table 1 Summary data for samples

Sample	Lat. (S)	Long. (E)	Description of locality	Sample depth (m)	Radioactivity data			D_{cosmic} (Gy kyr ⁻¹)	Dose rate (Gy kyr ⁻¹)	Palaeodose (Gy)	Age (kyr)
					K ₂ O (%)	Th (p.p.m.)	U (p.p.m.)				
Northeastern Kalahari											
890/2	18° 45'	25° 45'	Linear dune	2.0	0.08 ± 0.01	1.4 ± 0.10	0.7 ± 0.06	0.14	0.48 ± 0.05	10.5 ± 1.1	22 ± 3
891/1	18° 50'	25° 50'	Linear dune	1.8	0.17 ± 0.02	1.3 ± 0.10	0.5 ± 0.06	0.18	0.53 ± 0.05	13.1 ± 1.5	25 ± 4
891/2	18° 50'	25° 50'	Linear dune	0.9	0.17 ± 0.02	1.2 ± 0.10	0.5 ± 0.06	0.17	0.51 ± 0.05	12.6 ± 1.2	25 ± 3
892/1	19° 15'	26° 22'	Linear dune	1.2	0.07 ± 0.03	1.7 ± 0.10	0.6 ± 0.07	0.18	0.51 ± 0.09	11.0 ± 1.4	22 ± 5
892/2	19° 15'	26° 22'	Linear dune	0.7	0.07 ± 0.02	1.6 ± 0.10	0.6 ± 0.07	0.18	0.50 ± 0.08	7.0 ± 1.4	14 ± 4
894/1	18° 55'	26° 41'	Superficial dune mound	1.2	0.07 ± 0.03	2.5 ± 0.10	1.0 ± 0.08	0.19	0.67 ± 0.16	7.1 ± 0.4	10.6 ± 3
895/1	18° 43'	26° 57'	Minor linear dune	1.0	0.10 ± 0.03	1.5 ± 0.10	2.3 ± 0.10	0.18	0.91 ± 0.17	13.3 ± 3.0	14.6 ± 4
896/1	18° 00'	25° 50'	Linear dune	1.6	0.27 ± 0.05	2.6 ± 0.10	0.7 ± 0.08	0.18	0.75 ± 0.09	15.7 ± 5.6	21 ± 8
950/1	17° 58'	25° 38'	Linear dune	1.25	0.03 ± 0.001	1.8 ± 0.17	0.6 ± 0.13	0.17	0.47 ± 0.05	17.5 ± 3.4	37 ± 8
951/1	18° 02'	25° 41'	Linear dune	1.23	0.05 ± 0.002	1.8 ± 0.17	0.7 ± 0.14	0.17	0.51 ± 0.05	16.1 ± 2.3	32 ± 6
952/1	17° 55'	25° 28'	Extensively degraded dune mound	4.5	0.03 ± 0.001	1.8 ± 0.17	0.7 ± 0.13	0.17	0.49 ± 0.05	37.8 ± 4.6	77 ± 12
952/2	17° 55'	25° 28'	Extensively degraded dune mound	3.0	0.03 ± 0.001	1.9 ± 0.18	0.6 ± 0.13	0.17	0.47 ± 0.05	24.6 ± 2.9	52 ± 8
952/3	17° 55'	25° 28'	Extensively degraded dune mound	1.8	0.03 ± 0.001	1.5 ± 0.17	0.5 ± 0.13	0.18	0.43 ± 0.05	71 ± 11	164 ± 32
1003/1	18° 35'	25° 54'	Linear dune	1.0	0.41 ± 0.02	2.3 ± 0.21	0.8 ± 0.13	0.17	0.85 ± 0.09	12.7 ± 1.5	15 ± 2
1003/2	18° 35'	25° 54'	Linear dune	2.0	0.41 ± 0.02	2.8 ± 0.22	1.1 ± 0.15	0.17	0.96 ± 0.09	22.0 ± 1.9	23 ± 3
1004/A	19° 23'	26° 45'	Linear dune	0.5	0.17 ± 0.01	1.0 ± 0.15	0.5 ± 0.10	0.21	0.54 ± 0.06	9.6 ± 2.5	18 ± 5
1004/B	19° 23'	26° 45'	Linear dune	1.0	0.2 ± 0.01	1.3 ± 0.18	0.6 ± 0.10	0.18	0.57 ± 0.06	13.0 ± 2.5	23 ± 5
1004/C	19° 23'	26° 45'	Linear dune	1.5	0.2 ± 0.01	1.4 ± 0.24	0.7 ± 0.12	0.16	0.58 ± 0.07	16.2 ± 2.5	28 ± 5
1004/D	19° 23'	26° 45'	Linear dune	2.0	0.17 ± 0.01	1.6 ± 0.24	0.5 ± 0.11	0.15	0.52 ± 0.07	22.2 ± 3.3	43 ± 9
1004/E	19° 23'	26° 45'	Linear dune	2.5	0.17 ± 0.01	1.5 ± 0.24	0.8 ± 0.12	0.14	0.57 ± 0.07	24.3 ± 3.5	43 ± 8
1004/F	19° 23'	26° 45'	Linear dune	3.0	0.18 ± 0.01	1.6 ± 0.22	0.7 ± 0.12	0.13	0.55 ± 0.07	25.7 ± 2.8	46 ± 8
1004/G	18° 23'	26° 45'	Linear dune	3.5	0.19 ± 0.01	1.6 ± 0.16	0.7 ± 0.10	0.13	0.56 ± 0.05	21.8 ± 1.8	39 ± 5
1004/H	19° 23'	26° 45'	Linear dune	4.0	0.16 ± 0.01	1.6 ± 0.18	0.7 ± 0.09	0.12	0.53 ± 0.05	28.0 ± 2.9	53 ± 8
1004/I	19° 23'	26° 45'	Linear dune	4.5	0.29 ± 0.01	1.5 ± 0.17	0.6 ± 0.09	0.11	0.59 ± 0.07	26.7 ± 3.1	46 ± 7
1004/J	19° 23'	26° 45'	Linear dune	5.0	0.16 ± 0.01	1.9 ± 0.22	0.6 ± 0.11	0.10	0.51 ± 0.06	52.4 ± 5.5	103 ± 17
1004/K	19° 23'	26° 45'	Linear dune	5.5	0.17 ± 0.01	1.6 ± 0.17	0.6 ± 0.10	0.10	0.49 ± 0.06	54.7 ± 5.9	112 ± 18
1004/L	19° 23'	26° 45'	Linear dune	6.0	0.21 ± 0.01	1.7 ± 0.17	0.6 ± 0.10	0.09	0.52 ± 0.06	53.4 ± 6.5	102 ± 18
1005/A	15° 51'	26° 55'	Linear dune	0.5	0.08 ± 0.003	1.2 ± 0.11	0.5 ± 0.07	0.20	0.46 ± 0.03	4.3 ± 0.8	9 ± 2
1005/B	15° 51'	26° 55'	Linear dune	1.0	0.03 ± 0.003	1.5 ± 0.15	0.7 ± 0.07	0.18	0.50 ± 0.06	14.1 ± 2.2	28 ± 6
1005/C	15° 51'	26° 55'	Linear dune	1.5	0.08 ± 0.003	1.6 ± 0.19	0.7 ± 0.11	0.16	0.50 ± 0.05	10.1 ± 2.1	20 ± 5
1005/D	15° 51'	26° 55'	Linear dune	2.0	0.08 ± 0.003	1.3 ± 0.19	0.7 ± 0.13	0.14	0.46 ± 0.05	19.4 ± 2.0	42 ± 7
1005/E	15° 51'	26° 55'	Linear dune	2.5	0.09 ± 0.004	1.5 ± 0.18	1.0 ± 0.12	0.14	0.56 ± 0.05	16.6 ± 1.8	30 ± 4
1005/F	15° 51'	26° 55'	Linear dune	3.0	0.09 ± 0.004	1.5 ± 0.16	0.7 ± 0.11	0.13	0.48 ± 0.05	18.9 ± 2.1	40 ± 6
Southwestern Kalahari											
942/1	28° 24'	21° 30'	Linear dune	3.5	0.08 ± 0.02	1.7 ± 0.18	0.4 ± 0.12	0.13	0.79 ± 0.15	11.0 ± 1.0	14 ± 3
942/2	28° 24'	21° 30'	Linear dune	2.5	0.58 ± 0.02	2.0 ± 0.20	0.5 ± 0.13	0.14	0.86 ± 0.15	12.9 ± 0.7	15 ± 3
942/3	28° 24'	21° 30'	Linear dune	1.5	0.66 ± 0.03	2.5 ± 0.20	0.5 ± 0.14	0.16	0.98 ± 0.18	11.2 ± 0.9	12 ± 2
943/1	26° 33'	20° 35'	Linear dune	1.5	0.88 ± 0.04	1.0 ± 0.16	0.3 ± 0.12	0.17	1.00 ± 0.27	10.4 ± 2.1	10 ± 4
944/1	26° 39'	20° 36'	Linear dune	1.55	0.94 ± 0.04	1.5 ± 0.17	0.4 ± 0.12	0.14	1.08 ± 0.23	18.1 ± 2.6	17 ± 4
945/1	26° 14'	20° 36'	Basal linear dune field sand	1.75	0.78 ± 0.03	1.8 ± 0.18	0.5 ± 0.13	0.16	1.02 ± 0.18	23.5 ± 3.6	23 ± 5
946/1	26° 27'	21° 48'	Recticulate dune overlying linear	1.8	0.73 ± 0.03	1.4 ± 0.16	0.5 ± 0.13	0.15	0.94 ± 0.17	5.2 ± 1.5	6 ± 2
946b/1	26° 27'	20° 48'	Linear dune	1.5	0.78 ± 0.03	1.8 ± 0.17	0.4 ± 0.12	0.17	1.01 ± 0.20	26.8 ± 5.3	27 ± 7
947/1	26° 07'	20° 39'	Hummocky dune patch	1.0	0.36 ± 0.01	0.9 ± 0.16	0.2 ± 0.10	0.17	0.56 ± 0.16	1.0 ± 10.3	1.7 ± 0.7
947/2	26° 07'	20° 39'	Hummocky dune patch	1.0	0.51 ± 0.02	0.9 ± 0.15	0.3 ± 0.11	0.17	0.71 ± 0.16	1.0 ± 0.3	1.4 ± 0.5
948/1	26° 31'	20° 36'	Basal linear dune field sand	1.0	1.09 ± 0.04	2.6 ± 0.23	0.6 ± 0.15	0.19	1.37 ± 0.24	37.9 ± 6.4	28 ± 7
949/2	26° 41'	20° 10'	Lunette	1.0	1.17 ± 0.05	2.8 ± 0.31	0.8 ± 0.20	0.18	1.49 ± 0.27	2.1 ± 0.1	1.4 ± 0.3
949/3	26° 41'	20° 10'	Lunette	1.0	1.11 ± 0.04	3.2 ± 0.23	1.0 ± 0.19	0.22	1.56 ± 0.20	2.1 ± 0.1	1.4 ± 0.2
949/4	26° 41'	20° 10'	Lunette	1.0	1.39 ± 0.06	2.6 ± 0.23	0.7 ± 0.16	0.19	1.63 ± 0.27	1.8 ± 0.1	1.1 ± 0.2
949/5	26° 41'	20° 10'	Lunette	1.0	1.25 ± 0.05	2.5 ± 0.24	0.8 ± 0.18	0.23	1.58 ± 0.25	1.7 ± 0.1	1.1 ± 0.2

Samples for optical dating were collected by hammering 500 cm³ light-proof PVC cylinders horizontally into the vertical walls of freshly cleaned exposures. All samples were processed under subdued red light. Each was wet-sieved (retaining 90–125 µm size fraction) and immersed for 2 d in 1N HCl, followed by 2 d immersion in H₂O₂. Heavy minerals (density >2.72 g cm⁻³) were separated by magnetic and heavy-liquid (sodium polytungstate) methods, the remaining light minerals etched with 48% HF (60 min) and 35% H₂SiF₆ (4 d), and re-sieving at 75 µm. The quartz separates were mounted as monolayers (~5 mg per disk) onto 10-mm-diameter stainless steel disks using silicone spray adhesive. Palaeodoses were calculated using the multiple aliquot additive dose method²⁷. Aliquots were exposed to an argon laser (Coherent 2W, $\lambda = 514.5$ nm, power at sample ~40 mW cm⁻²), optically stimulated luminescence (OSL) emissions detected using a Philips 9635Q photomultiplier (PMT) filtered by BG-39 and Corning 7-51 glass filters. Aliquots were pre-heated either at 220 °C for 5 min (949/2–949/5), 160 °C for 16 h, or both (942/1–3). Growth curve data was via either natural or equal total dose normalization²⁸. Palaeodose estimates quoted are based on OSL from the full integrated laser exposure (~3J cm⁻²). Sample splits for dose rate determinations were crushed and homogenized by ring milling for 1 h. K⁴⁰, U and Th concentrations were estimated via neutron activation analysis and converted to dose rates²⁷. Saturation moisture content (W) was estimated at 0.4, and the estimated average moisture content over the burial period (F) as 0.1. (Here $W = (\text{wt H}_2\text{O})/(\text{wt sediment})$ and $F = (\text{moisture content})/W$). The cosmic-ray dose rate contribution (D_{cosmic} , in grays per kyr) was estimated via *in situ* γ -ray spectroscopy. The calculated dates incorporate random and systematic error²⁷, and are quoted to ± 1 standard deviation.

3.75° longitude, 19 vertical layers) was forced with observed SSTs for the period 1903–94. Empirical orthogonal functions (EOFs) of model austral summer (January–March) rainfall were calculated for the African land area and adjoining oceans using the ensemble average of four runs. An east–west dipole of rainfall, with centres over western southern Africa and the western Indian Ocean, captures 82% of rainfall variance of this broad region. The time coefficients of this EOF were correlated with global SSTs. The field shows a large coherent region of significant correlation in the southeastern Atlantic such that wet (dry) Kalahari conditions are associated with dry (wet) western Indian Ocean conditions, and positive (negative) SST anomalies in the southeastern Atlantic. Model circulation fields show the strengthening (weakening) of the South Atlantic anticyclone during Kalahari dry phases.

Evidence of moisture regimes since the last interglacial over

southern Africa and the southern Indian Ocean are scarce, although available palaeodata provide some insights into changing moisture balances over the past ~36 kyr before present^{6,23}. These studies confirm shifts between relatively wet and dry conditions on a 5–20 kyr timescale during this period (including the Holocene). The rainfall shift associated with the east–west dipole is consistent both with the Community Climate Model GCM experiments for 9 kyr ago and the present²⁴, and with elements of analogue models of southern African rainfall changes based on contemporary wet and dry spells^{6,7}. Kalahari dry (wet) phases are controlled by variability in the SST gradient across southern Africa such that large (weak) SST gradients lead to stationary easterly disturbances east (over) of the African landmass and eastward encroachment (westward retreat) of the South Atlantic anticyclone. SST estimates from ocean core data²⁵ reveals that since the last interglacial (marine

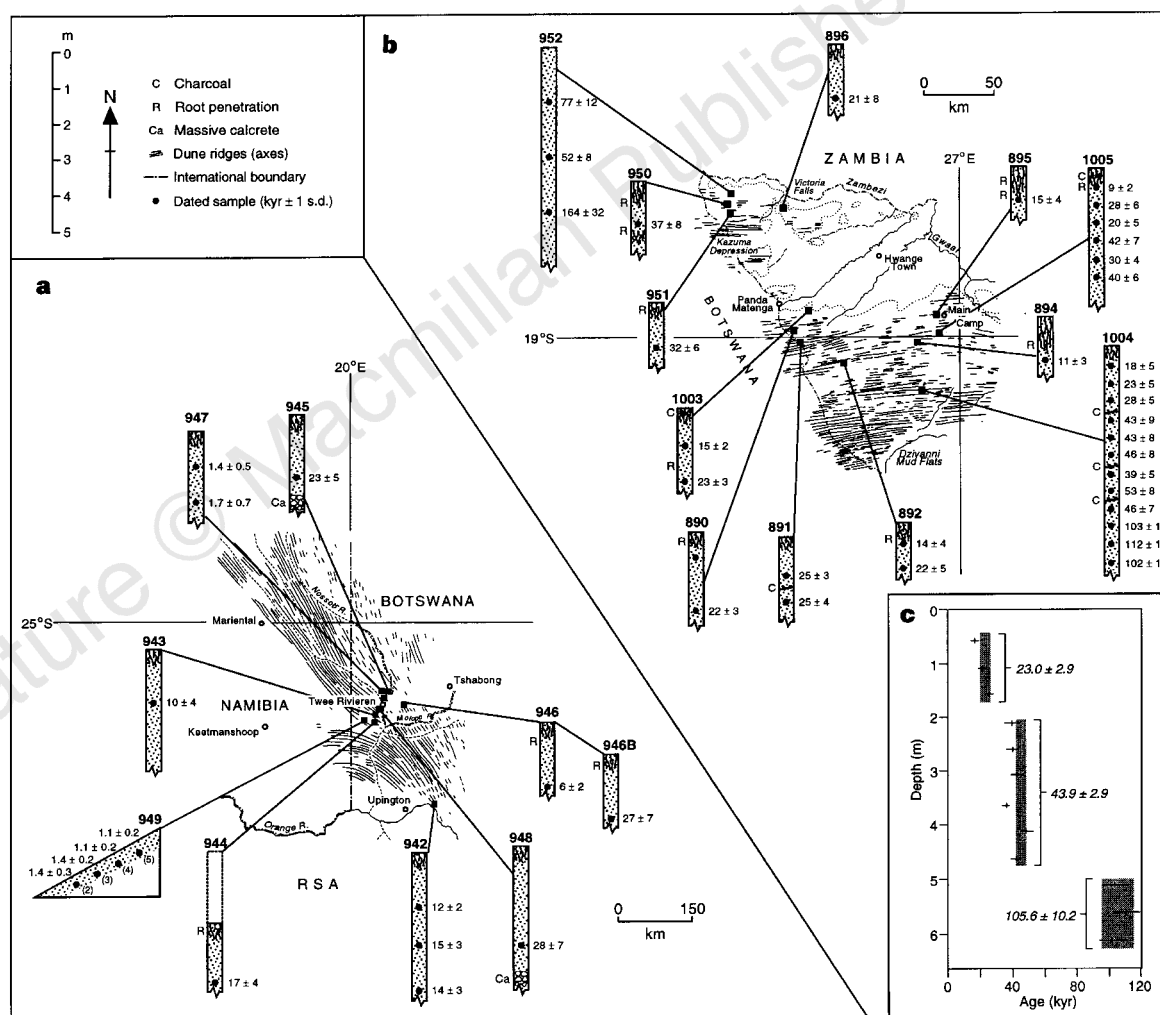


Figure 2 Summary stratigraphic and chronological data for dune sites: **a**, southwestern Kalahari; **b**, northeastern Kalahari; **c**, summary of optical dates and inferred aeolian depositional phases for site 1004 from the northeastern Kalahari (weighted average and standard deviations for grouped samples are provided). Note that different scales are used in **a** and **b**. Samples collected from degraded dune forms near basal massive calcretes at sites 945 and 948 provide the oldest ages in the southwestern Kalahari, in addition to the age from site 946B which was derived from a relatively weathered, fines-rich, linear dune. Analyses from sites 942 and 944 describe the main period of latest Pleistocene linear dune construction (~10–20 kyr ago) in the southwestern Kalahari which may have persisted into the Holocene (compare samples 943/1 and 946/1). The weighted

average for site 942, where three statistically similar samples were collected in stratigraphic order, is 13.2 ± 1.5 kyr (± 1 s.d.). Latest Holocene aeolian activity in the southwest is reflected in localized reactivation of hummocky dune forms (site 947) and lunette development (site 949). Age assessments for in the northeast are generally older than those from southwest. We use the record from site 1004 (**c**) as a master chronology for the area, from which three statistically distinct periods of dune activity are inferred spanning (at 1 s.d.) 95–115, 41–46 and 20–26 kyr ago. Latest Pleistocene dates younger than 20 kyr derived from site 1003, 892 and 895 are not statistically significantly different from the most recent phase of activity from site 1004.

isotope stage 5e) Indian Ocean variance is low (for example, core RC17-98, 13.22° S, 65.60° E; 0.18 °C) whereas that for the Atlantic is high (for example, core RC13-228, 22.33° S, 11.20° E; 6.1 °C). Recent analyses of southeastern Atlantic marine sediments have identified repeated short-lived (~5–10 kyr) periods of low SSTs²⁶, three of which correspond to periods of enhanced aridity in the Mega Kalahari. We propose that changes to the SST gradient and hence the northeast–southwest rainfall gradient are likely to originate in the Atlantic.

We have shown that the Mega Kalahari sedimentary system has exhibited a complex response to changes in Quaternary climates. Similar complexity can be expected in other low-latitude areas where extensive dune systems are preserved, which could provide critical information for GCM validation and inter-regional comparison; but direct chronologies of the development of such areas have not yet been established. □

Received 12 September 1996; accepted 23 May 1997.

1. Sarnthein, M. Sand deserts during glacial maximum and climatic optimum. *Nature* **272**, 396–398 (1978).
2. Sarnthein, M. & Koopman, B. Late Quaternary deep-sea record on northwest African dust supply and wind circulation. *Palaeoecol. Afr.* **12**, 238–253 (1980).
3. Street-Perrott, F. A., Marchand, D. S., Roberts, N. & Harrison, S. P. *Global Lake Level Variations from 18,000 to 0 Years Ago: A Palaeoclimatic Analysis* (Tech. Rep. TR046, US Dept of Energy, Washington DC, 1989).
4. de Menocal, P. B., Ruddiman, W. F. & Pokras, E. M. Influences of high- and low-latitude processes on African terrestrial climate: Pleistocene eolian records from equatorial Atlantic ocean drilling program Site 663. *Palaeoceanography* **8**, 209–242 (1993).
5. Thomas, D. S. G. & Shaw, P. A. *The Kalahari Environment* (Cambridge Univ. Press, 1991).
6. Cohen, A. L. & Tyson, P. D. Sea-surface temperature fluctuations during the Holocene off the south coast of Africa: implications for terrestrial climate and rainfall. *Holocene* **5**, 304–312 (1995).
7. Cockcroft, M. J., Wilkinson, M. J. & Tyson, P. D. The application of a present day climatic model to the late Quaternary in Southern-Africa. *Clim. Change* **10**, 161–181 (1987).
8. Van Zinderen Bakker, E. M. Comparison of late Quaternary climate change evolutions in the Sahara and the Namib-Kalahari region. *Palaeoecol. Afr.* **12**, 381–394 (1980).
9. Huntley, D. L., Godfrey-Smith, D. I. & Thewalt, M. L. W. Optical dating of sediment. *Nature* **313**, 257–267 (1985).
10. Henderson-Sellers, A. & Hansen, A.-M. *Climate Change Atlas: Greenhouse Simulations from Model Evaluation for Climate Assessment* (Kluwer, Dordrecht, 1995).
11. Schneider, S. H. (ed.) *Encyclopedia of Climate and Weather* (Oxford Univ. Press, New York, 1996).
12. Livingstone, I. & Thomas, D. S. G. in *The Dynamics and Environmental Context of Aeolian Sedimentary Systems* (ed. Pye, K.) 91–103 (Geol. Soc. London, 1993).
13. Wiggs, G. F. S., Thomas, D. S. G., Bullard, J. E. & Livingstone, I. Dune mobility and vegetation cover in the southwest Kalahari desert. *Earth Surf. Proc. Landf.* **20**, 515–530 (1995).
14. Thackeray, J. F. & Lee Thorpe, J. A. Isotopic analysis of equid teeth from Wonderwerk Cave, Northern Cape Province, South Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **99**, 141–150 (1992).
15. Partridge, T. C. *et al.* Late Pleistocene and Holocene climatic change in Southern Africa. *S. Afr. J. Sci.* **86**, 302–306 (1990).
16. Shaw, P. A. & Thomas, D. S. G. The Quaternary palaeoenvironmental history of the Kalahari, Southern Africa. *J. Arid Environ.* **32**, 9–22 (1996).
17. Brookes, G. A., Cowart, J. B. & Marais, E. Wet and dry periods in the southern African summer rainfall zone during the last 300 kyr from speleothem, tufa and sand dune age data. *Z. Fur. Geom.* (in the press).
18. Bloemendal, J. & de Menocal, P. Evidence for a change in the periodicity of tropical climate cycles at 2.4 Myr from whole-core magnetic susceptibility measurements. *Nature* **342**, 897–900 (1989).
19. Street-Perrott, F. A. & Perrott, R. A. Abrupt climate fluctuations in the tropics: the influence of Atlantic Ocean circulation. *Nature* **343**, 607–612 (1990).
20. de Menocal, P. & Rind, D. Sensitivity of Asian and African climate to variations in seasonal insolation, glacial ice cover, sea surface temperature, and Asian orography. *J. Geophys. Res.* **98**, 7267–7287 (1993).
21. Avery, D. M. Late Pleistocene coastal environment of the southern Cape Province of South Africa; micromammals from Klasies River mouth. *J. Archaeol. Sci.* **14**, 405–412 (1987).
22. Miller, D. F., Yates, R. J., Parkington, J. E. & Vogel, J. C. Radiocarbon-dated evidence relating to a mid-Holocene relative high sea-level on the south-western Cape Coast, South Africa. *S. Afr. J. Sci.* **89**, 35–44 (1993).
23. Sifeddine, A. *et al.* La sédimentation organique lacustre en zone tropicale sud au cours des 36000 dernières années (Lac Tririvakely, Madagascar). *C. R. Acad. Sci., Ser. IIa* **321**, 385–391 (1995).
24. Kutzbach, J. E., Guetter, P. J., Behling, P. J. & Selin, R. in *Global Climates Since the Last Glacial Maximum* (eds Wright, H. E. *et al.*) 24–93 (Univ. Minnesota Press, 1993).
25. Imbrie, J. & Duffy, A. *SPECMAP Archive* (NOAA Palaeoclimatology Program, Natl Geophys. Data Center, Boulder, CO, 1993).
26. Little, M. G. *et al.* Trade wind forcing of upwelling, seasonality, and Heinrich events as response to sub-Milankovitch climate variability. *Palaeoceanography* (in the press).
27. Aitken, M. J. *Thermoluminescence Dating* (Academic, London, 1985).
28. Stokes, S. Optical dating of selected late Quaternary aeolian sediment from the south-western United States. Thesis, Oxford Univ. (1994).

Acknowledgements. We thank P. Shaw and G. Haynes for discussions; M. Malifa, M. Moore, B. Adder, P. O'Connor, G. Wiggs, E. and K. Walker for technical and field assistance; P. Hayward for drafting assistance; G. A. Brookes, M. Little and D. Kroon for providing access to unpublished data; M. J. Aitken, A. K. Singhvi and C. K. Folland for advice on wider aspects of the research programme; and D. Sexton for help with model data. This work forms part of a wider project on southern African palaeoenvironments funded in part by the Trapnell Fund, the Universities of Oxford and Sheffield; R. Washington was supported by the DOE.

Correspondence and requests for materials should be addressed to S.S. (e-mail: stephen.stokes@geography.ox.ac.uk).

Sensitivity of earthquake cycles on the San Andreas fault to small changes in regional compression

Chi-yuen Wang* & Yongen Cai†

* Department of Geology & Geophysics, University of California, Berkeley, California 94720, USA

† Department of Geology, Peking University, Beijing 100871, China

The current pattern of slip^{1,2} within the San Andreas fault system in the San Francisco Bay area is distinctly different from the long-term slip pattern inferred from the geological record^{3,4}. This difference is not surprising because geological data record the accumulated displacements over many earthquake cycles, whereas geodetic data reveal the present-day slip pattern. It is not known, however, what mechanism triggers the change from the 'inter-seismic' slip pattern (when the San Andreas fault is locked) to the 'co-seismic' slip pattern (when the San Andreas fault ruptures in earthquake slip). Here we use numerical simulations of the entire seismic cycle on this complex fault system to show that the San Andreas fault may be in a critical state and sensitive to small perturbations in regional compression. In particular, we find that small increases in regional compression may lock the San Andreas fault, whereas small decreases in regional compression may release the locked segment and so permit co-seismic slip. This sensitivity suggests that cyclic changes in the regional stress field resulting from plate convergence and thrust faulting in the Coast Ranges could trigger major earthquakes on the San Andreas fault.

In the past decade, it has become clear that a regional compression subnormal to the San Andreas fault system (Fig. 1a) occurs in coastal California^{5,6}. This compression may be due to a small oblique convergence between the Pacific and the North American plates since 3.4–3.9 Myr ago⁷, producing thrust faulting^{8–11} (Fig. 1b) and thrust earthquakes¹² in the Coast Ranges. Several plate-motion models^{13–15} also show the presence of a small fault-normal component in the 'San Andreas discrepancy'. The average magnitude of this compression, inferred from stress orientation and surface heat flow, is⁵ ~100 MPa. Continuing, oblique plate convergence⁷ may cause the compression to increase slowly, whereas thrust faulting in the Coast Ranges may cause it to decrease intermittently; the magnitude of compression will therefore fluctuate with time (Fig. 1c).

The relative motion between the Pacific and the North-American plates drives the deformation in coastal California. A finite-element method is used to simulate this process. The model is a map-view layer of the seismogenic crust in central California¹⁶; it extends from Parkfield in the south to Point Arena in the north, and from the Pacific basin in the west to the Sierra Nevada foothills in the east. As the lateral dimensions of the model (350 km in the east–west direction and 450 km in the north–south direction) are much greater than the thickness of the seismogenic layer (10–15 km), a thin-plate approximation may be adopted. The base of this plate coincides with a master detachment beneath the San Francisco Bay area¹⁷. Calculation based on the crustal structure of central California¹⁸, the geotherm of the San Francisco Bay area¹⁹, and the rheological properties of crustal minerals²⁰ suggests that the detachment may correspond to a depth of minimum strength. Furthermore, pore pressure in many exploration wells in the region is nearly lithostatic²¹; thus the effective vertical stress may be small and a plane stress analysis may be appropriate. The geographical locations

of the major faults (Fig. 2) are taken from existing fault maps^{22,23}. Slip on the faults is assumed to obey Coulomb's friction law, with an effective frictional coefficient of 0.2 (refs 5, 6, 24). The upper-crustal blocks are assumed to behave as elastoplastic solid^{16,25}. Viscous relaxation of the crust and faults²⁶ is not considered.

As boundary conditions, a constant displacement rate of 39 mm yr^{-1} (ref. 16) along $\text{N } 30^\circ \text{ W}$ is imposed on the Pacific model boundary. At the same time, an effective compression of 100 MPa (ref. 5) is applied normal to this boundary. South of Hollister, the San Andreas system converges to a single strand. At latitudes of $36\text{--}37^\circ \text{ N}$, the fault creeps at a constant rate of 33 mm yr^{-1} (ref. 2). A displacement boundary condition of 33 mm yr^{-1} parallel to the San Andreas fault is thus applied on the southern boundary west of the San Andreas fault. This model¹⁶ correctly predicts the Quaternary fault slip rates (see also Table 1), the regional stress orientation, and the locations of crustal failure in the region; it is used as the starting model in this study.

We carried out four sets of numerical experiments. In the first experiment, two incremental steps, each with a shear displacement of $\Delta u = 39 \text{ m}$ (equivalent to time steps of 1,000 yr each), are applied on the Pacific model boundary; the normal force on the boundary is kept constant. The resulting slip pattern (not shown) in each step is identical to that in the starting model (Fig. 2a), demonstrating a quasi-steady fault-slip pattern under the imposed condition.

In the second experiment, an increase in shear displacement of $\Delta u = 39 \text{ m}$ (equivalent to a time step of 100 yr) and an increase in normal compression ΔP are applied on the Pacific boundary of the starting model. At $\Delta P = 0.1 \text{ MPa}$, the slip rate on the San Andreas fault (not shown) declines but remains the greatest in the area. At $\Delta P = 0.5 \text{ MPa}$, the fault segment between San Francisco and Hollister becomes locked (Fig. 2b), and the Hayward and the Calaveras faults become the fastest slipping faults in the system. We note the broad similarity between this slip pattern and the current inter-seismic slip pattern determined from geodetic measurements (Table 1). We also note that the magnitude of ΔP that is required to lock the San Andreas fault is similar to the stress drop in thrust earthquakes in the Coast Ranges²⁷. The stable transition from one experiment to another demonstrates the robustness of the solution.

The third experiment is identical to the second, except that ΔP and Δu are both increased by a factor of 10, while the ratio $\Delta P/\Delta u$ is kept constant. The resulting slip pattern (not shown) is identical to that in the previous experiment. This suggests that the ratio between ΔP and Δu , rather than their individual magnitudes, controls the slip pattern. Increasing the ratio tends to lock the San Andreas fault, whereas decreasing the ratio allows the fault to slip.

Last, we examine how the slip pattern may respond to a release of the regional compression, which may be caused by thrust faulting in the Coastal Ranges (Fig. 1c). We simulate this release by an abrupt reduction in compression at the Pacific model boundary of the inter-seismic model (Fig. 2b). The magnitude of this reduction (-0.5 MPa) is the same as that in the increase in compression required to lock the San Andreas fault. In response, the crustal blocks bounding the San Andreas fault abruptly accelerates in right-lateral motion (Fig. 2c), suggesting a transition from an inter-seismic state to a co-seismic slip; we note that during this event all the other faults in the system are locked. The maximum slip occurs on the San Andreas fault segment located on the San Francisco peninsula, where the fault was locked in the previous experiment. However, significant co-seismic slip occurs along the entire modelled San Andreas fault, extending both to the north and to the south, reaching the boundaries of the modelled domain.

The simulated inter-seismic (Fig. 2b) and co-seismic slip patterns (Fig. 2c) may be repeated with small increase/decrease in the regional compression. The long-term average of these repeated patterns is supposed to produce the geological slip pattern (Fig. 2a). The different responses of the San Andreas and the Hayward

Table 1 Observed and calculated slip rates at four crossings of the major faults of the San Andreas fault system

Faults	San Gregorio		San Andreas		Rodgers Creek/ Hayward		Green Valley/ Concord/ Calaveras	
	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
Inter-seismic								
North	—	—	0.6 ± 0.1	1	2 ± 2	3	5 ± 1	4
North-central	0	0	0	0	5 ± 1	3	3 ± 1	4
South-central	0	0	0	0	9 ± 1	7	3 ± 1	4
South	0	1	11 ± 3	$0\text{--}14^*$	—	—	12 ± 6	14
Geological								
North	—	—	24 ± 4	24	8 ± 2	8	6 ± 2	4
North-central	7 ± 1	4	19 ± 4	18	9 ± 1	8	6 ± 2	4
South-central	7 ± 1	4	19 ± 4	15	9 ± 1	8	8 ± 2	10
South	7 ± 1	4	22 ± 6	24	—	—	12 ± 2	10

All slip rates are given in mm yr^{-1} . The locations of the crossings within the San Francisco Bay area are shown in Fig. 2a.

*This crossing is located between the locked and the creeping segments on the San Andreas fault; the cited range represents slip rates in this transition.

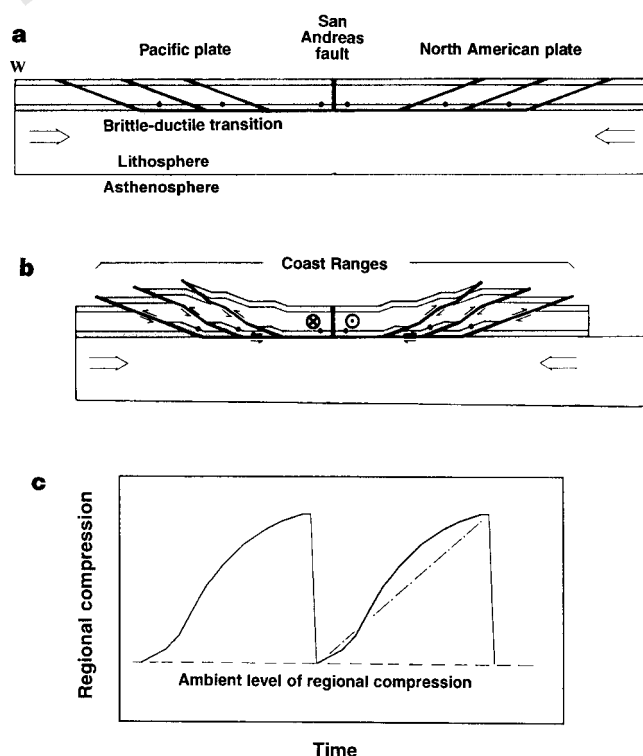
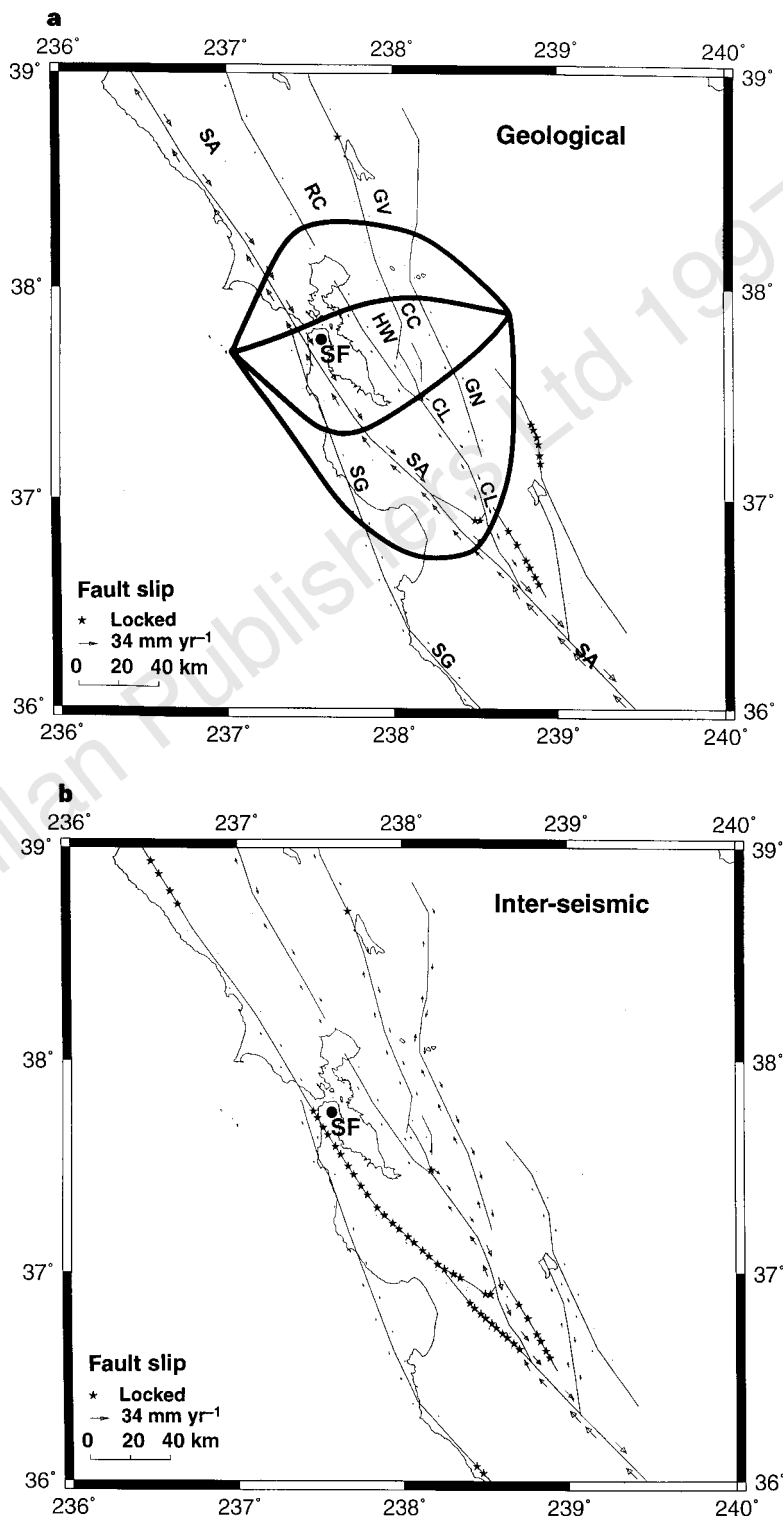


Figure 1 Schematic diagram illustrating fluctuations in regional compression in the San Francisco Bay area. **a**, Increase in the regional compression due to current oblique convergence between the Pacific and the North American plates (modified from ref. 10). The San Andreas fault is locked in this stage. **b**, Release in compression owing to thrust faulting in the Coast Ranges. In response, the crustal blocks bounding the San Andreas fault accelerate in right-lateral motion, as indicated by the symbols. **c**, Schematic diagram to show temporal fluctuations in the regional compression corresponding to the tectonic cycle presented in **a** and **b**. The amplitude of the fluctuations in compression are exaggerated; they may be minute relative to the ambient compression (see text). The increase in compression may be gradual and nonlinear with respect to time; the decrease in compression, on the other hand, may be abrupt.

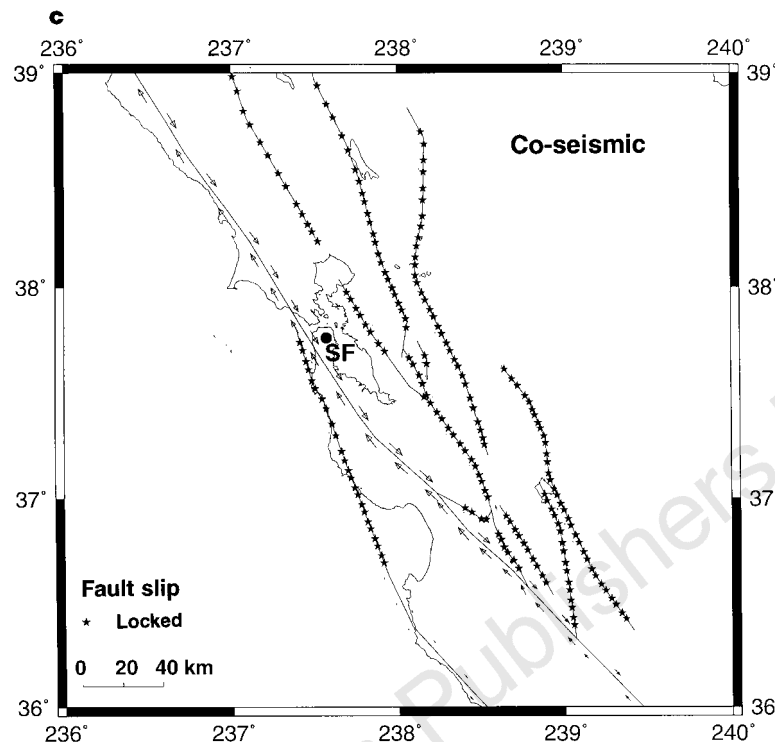
Figure 2 a, Slip pattern in the starting model¹⁶ for this study, which agrees closely with the Quaternary slip pattern on major faults in the region (Table 1). The four paths marked in thick lines showing crossings with major faults are, from the top downwards, the North path, the North-central path, the South-central path and the South path. CC, Concord fault; CL, Calaveras fault; GN, Greenville fault; GV, Green Valley fault; HW, Hayward fault; RC, Rodgers Creek fault; SA, San Andreas fault; SF, San Francisco; SG, San Gregorio fault. **b**, Slip pattern after applying $\Delta P = 0.5$ MPa and $\Delta u = 3.9$ m on the Pacific boundary of the starting model. A large segment of the San Andreas fault bordering the San Francisco Bay region becomes locked (indicated by the asterisks); the Hayward and the Calaveras become the fastest slipping faults in the system. This slip pattern is broadly similar to the current inter-seismic pattern determined from geodetic measurements (Table 1). **c**, Slip pattern after applying $\Delta P = -0.5$ MPa and $\Delta u = 0$ on the Pacific boundary of the inter-seismic model (**b**). The equivalent time step is $\Delta t = 0$, that is, decrease in compression takes place suddenly. In response, the crustal blocks bounding the San Andreas fault abruptly accelerate in right-lateral motion, suggesting co-seismic slip. Meanwhile, all the other faults in the system are locked.



faults to the changes in the regional compression may have a simple explanation. The average azimuth of the San Andreas fault on the San Francisco peninsula is a few degrees anticlockwise from the direction of the relative plate motion²⁸. Consequently, the regional compression has a component on the San Andreas fault in the direction opposing its right lateral motion. So when this component increases in magnitude with increasing regional compression, it may lock this segment of the San Andreas fault. On the other hand, the Hayward fault strikes $\sim N 30^\circ W$ and is thus nearly parallel to the plate-motion direction. Hence the fluctuations in the regional

compression would have little effect on the force component parallel to the fault.

Evidence in support of the suggestion that thrust faulting in the region may trigger earthquakes on the San Andreas fault exists in the historical earthquake record. Both the 1906 San Francisco and the 1989 Loma Prieta earthquakes on the San Andreas fault were preceded by several years of accelerated seismic moment releases in the Coast Ranges²⁹. In addition, immediately following the 1983 Coalinga thrust earthquake, which occurred in the Coast Ranges at a distance of 35 km from the San Andreas fault, accelerated slip



abruptly started on the segment of the San Andreas fault near Coalinga³⁰. But because this fault segment lies within the creeping section and little strain was accumulated², the induced accelerated slip was small.

Given that the crust is compressible, changes in the regional compression should be associated with changes in the lateral dimensions of the region. Thus the present model could, in principle, be tested against Global Positioning System (GPS) data from two or more well located stations. An order-of-magnitude estimate of the expected rate of change can be made as follows. We assume that a distance of 100 km separates two GPS stations located across the San Andreas fault system and the Coast Ranges. If the average magnitude of the regional changes in compression is 0.5 MPa (similar to the average stress-drop in thrust earthquakes) and the average linear compressibility of the upper crust between the two stations is $4 \times 10^{-6} \text{ MPa}^{-1}$ (similar to that of granodiorite³¹), a change of 0.2 m in distance between the two stations would be expected in one earthquake cycle. If, further, the inter-seismic interval is 200 yr and the rate of change in distance is constant, a rate of change of 1 mm yr^{-1} in distance between the two stations might occur during the inter-seismic interval. The considerable uncertainties in the many assumptions may cast doubt on the correctness of this estimate. But if the orders of magnitude of the assumed quantities are correct, the present model may be tested given ~ 10 yr of continuous GPS data across the Coast Ranges. □

Received 21 October 1996; accepted 3 June 1997.

1. Thatcher, W. in *The San Andreas Fault System, California*. Prof. Pap. US Geol. Surv. **1515**, 189–205 (1990).
2. Lisowski, M., Savage, J. C. & Prescott, W. H. The velocity field along the San Andreas fault in central and south California. *J. Geophys. Res.* **96**, 8369–8389 (1991).
3. Brown, R. D. Jr in *The San Andreas Fault System, California*. Prof. Pap. US Geol. Surv. **1515**, 83–113 (1990).
4. Kelson, K. I., Lettis, W. R. & Lisowski, M. in *Proc. 2nd Conf. Earthquake Hazards in the Eastern San Francisco Region* (eds Borchardt, G. et al.) 31–33 (Spec. Publ. 113, Calif. Division Mines & Geol. Sacramento, 1992).
5. Mount, V. S. & Suppe, J. State of stress near the San Andreas fault: Implications for wrench tectonics. *Geology* **15**, 1143–1146 (1987).
6. Zoback, M. D. et al. New evidence on the state of stress of the San Andreas fault system. *Science* **238**, 1105–1111 (1987).
7. Harbert, W. Late Neogene relative motions of the Pacific and North America plates. *Tectonics* **10**, 1–15 (1991).

8. Page, B. M. in *Proc. Conf. Earthquake Hazards in the Eastern San Francisco Bay Region* (eds Hart, E. W., Hirschfeld, S. E. & Schulz, S. S.) 1–10 (Calif. Division of Mines & Geology, Sacramento, 1982).
9. Aydin, A. in *Prof. Conf. Earthquake Hazards in the Eastern San Francisco Bay Region* (eds Hart, E. W., Hirschfeld, S. E. & Schulz, S. S.) 11–21 (Calif. Division Mines & Geol., Sacramento, 1982).
10. Namson, J. S. & Davis, T. L. Seismically active fold and thrust belt in the San Joaquin Valley, Central California. *Geol. Soc. Am. Bull.* **100**, 257–273 (1988).
11. Jones, D. L. et al. Neogene transpressive tectonics of the California Coast Ranges. *Tectonics* **13**, 561–574 (1994).
12. Ekstrom, G., Stein, R. S., Eaton, J. P. & Eberhart-Phillips, D. Seismicity and geometry of a 110-km-long blind thrust fault: 1. The 1985 Kettleman Hills, California, earthquake. *J. Geophys. Res.* **97**, 4843–4864 (1992).
13. Minster, J. B. & Jordan, T. H. Present-day plate motions. *J. Geophys. Res.* **83**, 5331–5353 (1978).
14. DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
15. Argus, D. F. & Gordon, R. G. Current Sierra Nevada–North American motion from very long Baseline Interferometry: implications for the kinematics of the western United States. *Geology* **19**, 1085–1088 (1991).
16. Wang, C. Y., Cai, Y. & Jones, D. L. Predicting crustal failure in the Greater San Francisco Bay area. *Geology* **23**, 771–774 (1995).
17. Brocher, T. M. et al. Seismic evidence for a lower-crustal detachment beneath San Francisco Bay, California. *Science* **265**, 1436–1439 (1994).
18. Fuis, G. S. & Mooney, W. D. in *The San Andreas Fault System, California*. Prof. Pap. US Geol. Surv. **1515**, 207–238 (1990).
19. Williams, C. F. & Galanis, P. S. Jr Heat-flow measurements in the vicinity of the Hayward Fault, California. (Open File Rep. 94-692, US Geol. Surv., Reston, 1994).
20. Kohlstedt, D. L., Evans, B. & Mackwell, S. J. Strength of the lithosphere: Constraints imposed by laboratory experiments. *J. Geophys. Res.* **100**, 17587–17602 (1995).
21. Berry, F. A. F. High fluid potentials in California Coast Ranges and their tectonic significance. *Bull. Am. Assoc. Petrol. Geol.* **57**, 1219–1248 (1973).
22. Jennings, C. W. Fault map of California with locations of volcanoes, thermal springs, and thermal wells. (Geol. Map 1, scale 1:750,000, Calif. Division Mines & Geol., Sacramento, 1975).
23. Graymer, R., Jones, D. L. & Brabb, E. E. (Open File Map 94-622, US Geol. Surv., Reston, 1994).
24. Lachenbruch, A. H. & Sass, J. H. Heat flow and energetics of the San Andreas fault zone. *J. Geophys. Res.* **85**, 6185–6222 (1980).
25. Braun, J. & Beaumont, C. Three-dimensional numerical experiments of strain partitioning at oblique plate boundaries: Implications for contrasting tectonic styles in the southern Coast Ranges, California, and central South Island, New Zealand. *J. Geophys. Res.* **100**, 18059–18074 (1995).
26. Thatcher, W. Nonlinear strain buildup and the earthquake cycle on the San Andreas fault. *J. Geophys. Res.* **88**, 5893–5902 (1983).
27. Scholz, C. H. *The Mechanics of Earthquake Faulting* (Cambridge Univ. Press, 1990).
28. Scholz, C. H. The Black Mountain asperity: seismic hazard of the southern San Francisco Peninsula, California. *Geophys. Res. Lett.* **12**, 717–719 (1985).
29. Jaume, S. C. & Sykes, L. R. Evolution of moderate seismicity in the San Francisco Bay region, 1850 to 1993; seismicity changes related to the occurrence of large and great earthquakes. *J. Geophys. Res.* **101**, 765–790 (1996).
30. Mavko, G. M., Schulz, S. & Brown, B. D. Effects of the 1983 Colinga, California earthquake on creep along the San Andreas fault. *Bull. Seismol. Soc. Am.* **75**, 475–489 (1985).
31. Birch, F. Handbook of Physical Constants. *Mem. Geol. Soc. Am.* **97**, 97–174 (1966).

Acknowledgements. We thank D. Jones for discussions, and E. Jones and C. Beaumont for comments.

Correspondence should be addressed to C.-y. W.

Plume-like neon in a metasomatic apatite from the Australian lithospheric mantle

Takuya Matsumoto^{*†}, Masahiko Honda^{*}, Ian McDougall^{*}, Igor Yatssevich^{*} & Suzanne Y. O'Reilly[‡]

^{*} Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia

[‡] Key Centre for Geochemical Evolution and Metallogeny of Continents (GEMOC), School of Earth Sciences, Macquarie University, Sydney, NSW 2109, Australia

The nature of metasomatizing fluids and melts in the mantle are of interest for understanding the chemical evolution of the Earth's interior^{1–3}. The study of noble-gas isotopes in appropriate mantle-derived samples has the potential to provide valuable insight into this question, by constraining the origin of the fluids and the timing of metasomatic events. Here we report the application of neon-isotope systematics to investigate the metasomatic history of apatite grains in spinel-lherzolite xenoliths from the Australian lithospheric mantle. We find that the apatite has a neon-isotope signature similar to that associated with plume-related volcanism, as is found in Hawaii, whereas coexisting mineral phases (olivine and amphibole) and non-apatite-bearing lherzolites have isotope signatures more typical of mid-ocean-ridge basalts. The occurrence of plume-like neon in the apatite implicates deep plume-like material beneath southeastern Australia as the source of the metasomatizing agent.

We have analysed noble-gas abundances and isotopic ratios in gases released from mineral separates from spinel lherzolites hosted in Pliocene to Recent alkali basalts of the Newer Volcanics, south-eastern Australia. Geothermometric studies indicate that these spinel lherzolites are derived from the uppermost lithospheric mantle, with the inferred depth of origin in the interval from the crust–mantle boundary (~30 km) to the spinel- to garnet-lherzolite transition (~60 km)⁴. One sample (BM901), from Bullenmerri maar, contains apatite and amphibole as accessory volatile-bearing (that is, containing Cl, F and/or OH) phases and these minerals were separated for analysis. Sample BM901 consists of olivine, 63%; clinopyroxene, 9%; orthopyroxene, 15%; spinel, 0.5%; amphibole, 13%; apatite, 0.5%, by mass. The other samples measured are olivine separates from lherzolites which are free of volatile-bearing minerals (hereafter referred to as 'dry lherzolites'). We will focus here on helium and neon isotopic compositions measured in these samples (Table 1); results on the heavier noble gases will be discussed elsewhere.

Neon measurements for all the olivine and amphibole separates (from BM901 and two dry lherzolites) define a linear array whose slope is similar to the well-established MORB-correlation line (Fig. 1a)⁵. The ³He/⁴He ratios observed in these samples are also within the range reported from MORBs ((12 ± 2) × 10^{−6}; ref. 6), consistent with the results obtained previously from other xenoliths from the same field⁷.

However, isotopic compositions of helium and neon observed in the apatite separated from BM901 appear to be distinct from MORB-like compositions. We found low ³He/⁴He ratios (~4 × 10^{−7}) in the apatite, so it is probably dominated by radiogenic ⁴He reflecting relatively high concentrations of U and Th in the apatite (70 and 150 p.p.m. respectively⁸). In addition, isotopic compositions of neon in the apatite indicate the presence of two

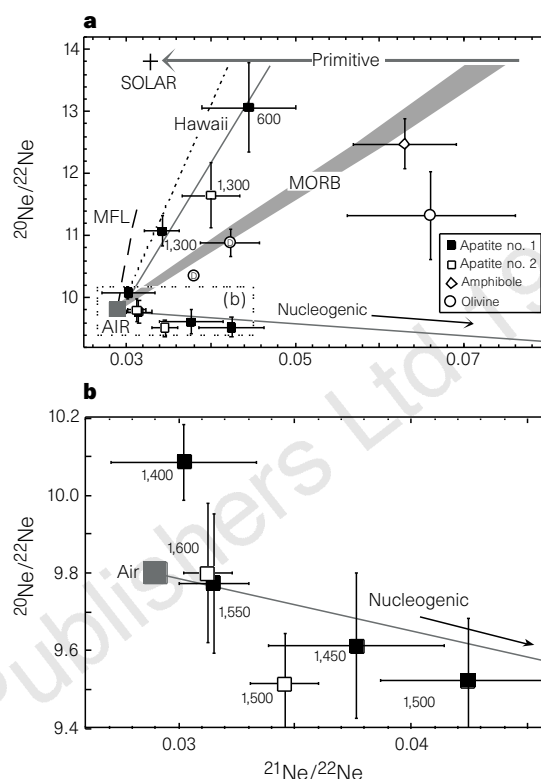


Figure 1 **a**, Three-isotope plot for neon extracted from apatite, amphibole and olivine separates from the spinel lherzolite BM901, Bullenmerri, Victoria, Australia, and from olivine separates from dry lherzolites from Mt Gambier, South Australia (shown as 'D'). **b**, Expanded-scale view of the area of **a** enclosed in the dotted box. The atmospheric and solar neon compositions are shown as 'AIR' and 'SOLAR', respectively. The mass-fractionation line (long-dashed) with initial atmospheric neon composition is labelled as 'MFL'. The MORB-correlation line (shaded area) is from Sarda *et al.*⁵. The short-dashed line labelled 'Hawaii' is from Honda *et al.*^{10,11} for basaltic samples from Loihi and Kilauea. The line labelled 'Primitive' indicates locus of compositions derived from a primordial solar composition augmented by nucleogenic neon production. The solid line with positive slope in **a** is the best-fit regression line derived from the isotopic composition of neon released between 600 and 1,400 °C from the apatite samples. The solid line labelled 'Nucleogenic' is a mixing line between the calculated *in situ* nucleogenic component in the apatite (0.0765 and 0.792 for ²⁰Ne/²²Ne and ²¹Ne/²²Ne, respectively) and atmospheric neon. Numbers shown near the apatite data refer to the temperature at which the samples were heated to release the noble gases.

distinct components, both of which are different from MORB-like compositions. As shown in Fig. 1a, neon results from two aliquots of the apatite (labelled no. 1 and no. 2), on gases released in temperature fractions from 600 to 1,400 °C, define a linear array extending from atmospheric neon to a component having higher ²⁰Ne/²²Ne and ²¹Ne/²²Ne ratios. A second component is released from the apatite at temperatures between 1,450 and 1,600 °C (Fig. 1b). The neon isotopic ratios observed in the higher-temperature fractions also define a linear array, indicating the presence of a component having lower ²⁰Ne/²²Ne and higher ²¹Ne/²²Ne ratios compared with atmospheric neon.

It is difficult to explain these linear trends by mass-dependent isotopic fractionation during the step-heating experiments, as the slopes of the correlation lines are very different from that calculated for mass fractionation of atmospheric neon (Fig. 1a). Also, the highest measured ²⁰Ne/²²Ne ratio is substantially higher than the upper limit expected from single-stage mass fractionation of the atmospheric composition (²⁰Ne/²²Ne < 10.3). Multiple stages of

[†] Present address: Department of Earth and Space Science, Graduate School of Science, Osaka University, Toyonaka, Osaka 560, Japan.

Table 1 He and Ne in spinel lherzolites

Temp. (°C)	^4He ($10^{-8} \text{ cm}^3 \text{ STP g}^{-1}$)	$^3\text{He}/^4\text{He}$ (10^{-6})	^{22}Ne ($10^{-12} \text{ cm}^3 \text{ STP g}^{-1}$)	$^{20}\text{Ne}/^{22}\text{Ne}$	$^{21}\text{Ne}/^{22}\text{Ne}$
BM901 olivine (1.1736 g)					
900	0.367 ± 0.004	b.d.	—	—	—
1,800	2.71 ± 0.03	14.3 ± 0.4	1.31 ± 0.19	11.33 ± 0.71	0.0661 ± 0.0091
Total	3.08 ± 0.03	14.3 ± 0.4	1.31 ± 0.19	11.33 ± 0.71	0.0661 ± 0.0091
BM901 amphibole (0.3400 g)					
600	0.285 ± 0.004	b.d.	1.09 ± 0.30	8.77 ± 0.68	0.0338 ± 0.0074
1,100	4.60 ± 0.05	12.0 ± 0.8	—	—	—
1,600	2.25 ± 0.02	14.3 ± 2.7	3.99 ± 0.50	12.48 ± 0.40	0.0629 ± 0.0061
Total	7.14 ± 0.05	13.6 ± 1.1	5.08 ± 0.59	11.68 ± 0.40	0.0566 ± 0.0052
BM901 apatite no. 1 (0.1491 g)					
350	—	—	—	—	—
600	90.0 ± 0.9	0.40 ± 0.10	3.99 ± 0.79	13.07 ± 0.72	0.0445 ± 0.0056
800	35.4 ± 0.4	b.d.	—	—	—
1,050	149 ± 1	0.44 ± 0.05	—	—	—
1,300	9.84 ± 0.10	b.d.	13.4 ± 1.4	11.08 ± 0.25	0.0343 ± 0.0021
1,400	0.024 ± 0.007	b.d.	24.4 ± 3.4	10.09 ± 0.10	0.0302 ± 0.0032
1,450	—	—	15.1 ± 2.1	9.61 ± 0.19	0.0377 ± 0.0038
1,500	—	—	12.0 ± 1.4	9.52 ± 0.16	0.0425 ± 0.0038
1,550	—	—	11.5 ± 0.9	9.77 ± 0.18	0.0315 ± 0.0015
1,600	—	—	—	—	—
Total	284 ± 2	0.42 ± 0.05	80.4 ± 4.6	10.18 ± 0.09	0.0350 ± 0.0014
BM901 apatite no. 2 (0.0891 g)					
600	112 ± 1	b.d.	—	—	—
1,300	423 ± 4	0.23 ± 0.03	15.9 ± 3.0	11.65 ± 0.52	0.0400 ± 0.0034
1,500	0.339 ± 0.013	b.d.	46.8 ± 4.3	9.51 ± 0.13	0.0346 ± 0.0015
1,600	0.054 ± 0.014	b.d.	29.2 ± 3.6	9.80 ± 0.18	0.0313 ± 0.0010
Total	535 ± 4	0.23 ± 0.03	91.9 ± 6.4	10.23 ± 0.29	0.0345 ± 0.0035
GAMVL2 olivine (2.9447 g)					
900	0.270 ± 0.004	b.d.	0.27 ± 0.05	10.68 ± 0.39	0.0309 ± 0.0030
1,800	9.42 ± 0.10	9.9 ± 0.3	0.52 ± 0.06	10.88 ± 0.22	0.0423 ± 0.0035
Total	9.69 ± 0.10	9.9 ± 0.3	0.79 ± 0.07	10.81 ± 0.20	0.0384 ± 0.0026
GAMVL11 olivine (1.9606 g)					
900	0.156 ± 0.006	b.d.	5.8 ± 0.2	9.85 ± 0.08	0.0286 ± 0.0005
1,800	34.2 ± 0.3	10.9 ± 0.3	8.8 ± 0.2	10.35 ± 0.05	0.0381 ± 0.0004
Total	34.4 ± 0.3	10.9 ± 0.3	14.5 ± 0.2	10.15 ± 0.05	0.0343 ± 0.0003

BM901 is an amphibole- and apatite-bearing metasomatized spinel lherzolite from a Quaternary maar, Lake Bullenmerri (38° 15' S, 143° 06' E), in western Victoria, Australia. GAMVL2 and GAMVL11 are spinel lherzolites (without volatile-bearing minerals, that is, dry peridotites) from Quaternary scoria, Mt Gambier (37° 49' S, 140° 45' E), in South Australia. Duplicate analyses were made for BM901 apatite (no. 1 and no. 2). Olivine and amphibole separates for noble-gas study were hand-picked from coarsely crushed whole-rock samples and washed in 7% HF for 10 min, and in acetone and ethanol for 30 min in an ultrasonic bath. For whole-rock and apatite samples, the HF treatment was not applied, as apatites are easily dissolved by acid treatment. Elemental and isotopic compositions of helium and neon were determined in a VG5400 mass spectrometer at the Australian National University. The procedures for noble-gas extraction, purification and mass analysis are essentially the same as those described in ref. 11. Quoted uncertainties (1 σ) include uncertainties in the sensitivity determined by repeated analysis of standard gases, uncertainties in the correction factors for mass discrimination, uncertainties in the blank correction and uncertainties in interference corrections in the case of neon. The detection limits for ^3He and ^{21}Ne are $5.0 \times 10^{-14} \text{ cm}^3 \text{ STP}$ and $2.5 \times 10^{-14} \text{ cm}^3 \text{ STP}$, respectively. Measurements below the detection limit are listed as 'b.d.'. The blank correction was applied to all results listed here, assuming the blank had an atmospheric isotopic composition. Typical blank levels of ^4He and ^{22}Ne during these experiments were $4 \times 10^{-11} \text{ cm}^3 \text{ STP}$ and $3 \times 10^{-13} \text{ cm}^3 \text{ STP}$, respectively. If the measured peak intensity of the reference isotope is 1 σ variation of the system blank, the measurement is regarded as a blank ('—'), and corresponding isotopic ratios are not shown.

mass fractionation would be needed to produce $^{20}\text{Ne}/^{22}\text{Ne}$ ratios as high as 13. We regard this as unrealistic for a number of reasons, including the absence of evidence for complementary isotopic compositions. We believe that the observed linear trends are not experimental artefacts, but are most easily understood as mixing lines, with atmospheric neon as a common end member. These neon signatures were confirmed by the duplicated step-heating measurements on the apatite, which yield essentially the same results. However, to gain a better understanding of the location of these two components in the apatite samples, it is obvious in retrospect that vacuum crushing should have been undertaken on at least one aliquot before step-heating. Insufficient remaining apatite separate prevented us from conducting a crushing experiment subsequently.

In general, correlated $^{20}\text{Ne}/^{22}\text{Ne}$ and $^{21}\text{Ne}/^{22}\text{Ne}$ ratios higher than atmospheric neon observed in mantle-derived samples are interpreted as representing a mixture of atmospheric neon and a mantle component, comprising primordial solar-like neon ($^{20}\text{Ne}/^{22}\text{Ne} = 13.8$, $^{21}\text{Ne}/^{22}\text{Ne} = 0.0328$)⁹ and nucleogenic neon. We suggest that neon in each mantle reservoir has a solar-like $^{20}\text{Ne}/^{22}\text{Ne}$ ratio and an elevated $^{21}\text{Ne}/^{22}\text{Ne}$ ratio reflecting the time-integrated U + Th/Ne ratio in the source^{10,11}. The increase in $^{21}\text{Ne}/^{22}\text{Ne}$ ratio is attributed to production of nucleogenic ^{21}Ne in the mantle; when U and Th decay, they produce α -particles which interact with O to produce ^{21}Ne . On this basis, the more primitive

mantle source should have a lower $^{21}\text{Ne}/^{22}\text{Ne}$ ratio, reflecting a relatively low time-integrated U + Th/Ne ratio.

The steeper slope of the neon correlation line through the apatite results, with respect to the MORB neon correlation line, indicates that the mantle source for the apatite has a neon signature more primitive than that of the MORB source. Such primitive neon signatures have been reported previously for plume-related samples from suboceanic environments, such as in oceanic basalts from Loihi and Kilauea, Hawaii^{5,10–12}, and ultramafic xenoliths from Réunion¹³, Kerguelen¹⁴ and Samoa¹⁵, indicating the presence of a primitive noble-gas reservoir in the suboceanic mantle, presumably located below the depleted upper mantle.

In continental areas, more primitive neon signatures compared with MORBs have thus far only been reported in fluids from the Yellowstone area, USA¹⁶. Together with the high $^3\text{He}/^4\text{He}$ ratio (2.2×10^{-5}) in the fluids, we take this as evidence for plume-like activity below Yellowstone. We are not aware of any previous report of a plume-like noble-gas signature in mantle-derived xenoliths from the subcontinental lithospheric mantle. But neon results from BM901 apatite suggest that a similar primitive noble-gas reservoir indeed exists in the Australian subcontinental mantle. This reservoir has a neon isotopic composition very similar to that found in the Hawaiian plume, currently the most primitive source known in terms of the neon isotopic ratios observed in mantle-derived samples (Fig. 1a).

It has become generally accepted that the southward-younging volcanism of the central volcanoes in eastern Australia over the past 35 Myr effectively records the northward movement of the Indo-Australian plate over a mantle hotspot¹⁷. The youthful basaltic activity (0–5 Myr) in the Newer Volcanics is consistent with the projected location of the hotspot. Furthermore, Newer Volcanics basalts have 'plume'-type chemical characteristics (see, for example, refs 18, 19). Western Victoria has the highest heat flow at present known in Australia^{20,21}, and tomographic images reveal that this area is correlated with pronounced low-velocity anomalies at depths between 80 and 140 km (ref. 22). Therefore, geochemical and geophysical evidence is consistent with the presence of hot plume material beneath southeastern Australia.

The observed plume-like neon signature in the apatite is apparently not a major component in the lithospheric mantle, as the olivine separates from the dry lherzolites yield MORB-like helium and neon isotopic ratios. If these are representative samples, then perhaps the majority of the subcontinental lithospheric mantle is characterized by helium and neon isotopic compositions similar to those observed in MORBs. As the apatite is regarded as a product of the metasomatic reaction between the mantle wall rock and fluids or melts in the lithospheric mantle, we suggest that the observed plume-like neon signature in the apatite is metasomatically introduced into the lithospheric mantle by the fluid or melt responsible for the apatite formation.

The precipitation of apatite rich in Cl, F and CO₂ indicates that the apatite-forming metasomatic agent is carbonatitic melt or carbonate-rich fluid of uncertain origin^{8,23–26}. It has been suggested that the observed 'bulk-Earth' Sr and Nd isotopic signatures in the metasomatic apatites may indicate their derivation from a less-depleted (plume-like) mantle source, or may simply represent the averaged isotopic composition acquired during movement of the metasomatizing fluid through an isotopically heterogeneous mantle²⁷. However, the present neon results strongly support the involvement of plume-like material in the apatite-forming event. As this signature is not observed in amphibole in the same lherzolite, the amphibole-forming metasomatizing agent is likely to have been derived from a different mantle source.

In contrast, as noted previously, the neon released from the apatites at high temperatures (1,450–1,600 °C) shows systematically lower ²⁰Ne/²²Ne and higher ²¹Ne/²²Ne ratios with respect to atmospheric neon. This signature can be attributed to either spallogenic or nucleogenic neon. Spallation by cosmic-ray-induced particles can be excluded, however, because of the absence of excess ³He expected from this process. Thus, nucleogenic reactions are probably responsible. Taking into account the chemical composition of the apatite (F, 0.26 wt%; ref. 8), the possible nuclear reactions are ¹⁹F(α,n)²²Na(β[−])²²Ne (ref. 28) and ¹⁹F(α,p)²²Ne (ref. 29), as well as ¹⁸O(α,n)²¹Ne (ref. 28). As shown in Fig. 1b, the high-temperature data plot close to a mixing line between atmospheric neon and the theoretically expected production ratio of nucleogenic neon in the apatite BM901 (²⁰Ne/²²Ne = 0.0765 and ²¹Ne/²²Ne = 0.792 on the basis of the known chemical compositions of this particular apatite⁸ and the algorithm developed by Yatsevich and Honda³⁰), suggesting that the neon composition observed in the higher-temperature fractions is actually an *in situ* nucleogenic component accumulated in the apatite. The accumulation time required for producing the observed amounts of nucleogenic ²¹Ne and ²²Ne is estimated to be ~1 Myr (for example, amounts of nucleogenic ²¹Ne and ²²Ne in the aliquot no. 1 are 3.4×10^{-13} and 6.6×10^{-13} cm³ STP g^{−1} and production rates are estimated to be 5.2×10^{-19} and 6.7×10^{-19} cm³ STP g^{−1} yr^{−1}, respectively). This inferred accumulation time is much greater than the suggested formation age of the Bullenmerri maar (which has a minimum age of ~25,000 yr before present, inferred from ¹⁴C dating carried out on lake sediments in the maar³¹), thus it is probably recording part of its mantle history.

The apatite seems to have preserved two distinct neon compo-

nents which did not isotopically equilibrate during mantle residence. The nucleogenic neon is probably located in the structure of the apatite, as the nuclear reactions target the elements in the crystal lattice. In contrast, the plume-like neon is trapped in somewhat less retentive sites, as this component is released at lower temperatures. The most plausible candidate for the site of the plume-like neon would be fluid inclusions that are abundant in the apatite, because decrepitation of the fluid inclusions would occur at temperatures below that of apatite melting (~1,600 °C). If this is the case, then the high confining pressures in the mantle would prevent decrepitation of the inclusions, and fluid–crystal partitioning would prevent diffusive loss of neon from the inclusions. □

Received 5 August 1996; accepted 20 May 1997.

- McDonough, W. F. & McCulloch, M. T. The southeast Australian lithospheric mantle: isotopic and geochemical constraints on its growth and evolution. *Earth Planet. Sci. Lett.* **86**, 327–340 (1987).
- Meen, K. J., Egger, D. H. & Ayers, J. C. Experimental evidence for very low solubility of rare-earth elements in CO₂-rich fluids at mantle conditions. *Nature* **340**, 301–303 (1989).
- O'Reilly, S. Y. & Griffin, W. L. Mantle-metasomatism beneath western Victoria, Australia: I. Metasomatic processes in Cr-diopside lherzolites. *Geochim. Cosmochim. Acta* **52**, 433–447 (1988).
- Griffin, W. L., Wass, S. Y. & Hollis, J. D. Ultramafic xenoliths from Bullenmerri and Gnotuk maars, Victoria, Australia: Petrology of sub-continental crust-mantle transition. *J. Petrol.* **25**, 53–87 (1984).
- Sarda, P., Staudacher, T. & Allègre, C. J. Neon isotopes in submarine basalts. *Earth Planet. Sci. Lett.* **91**, 73–88 (1988).
- Craig, H. & Lupton, J. E. in *The Sea* (ed. Emiliani, C.) 391–428 (Wiley, New York, 1981).
- Porcelli, D., O'Nions, R. K. & O'Reilly, S. Y. Helium and strontium isotopes in ultramafic xenoliths. *Chem. Geol.* **54**, 237–249 (1986).
- O'Reilly, S. Y., Griffin, W. L. & Ryan, C. G. Residence of trace elements in metasomatized spinel lherzolite xenoliths: a proton-microprobe study. *Contrib. Mineral. Petrol.* **109**, 93–113 (1991).
- Benkert, J.-P., Baur, H., Signer, P. & Wieler, R. He, Ne, and Ar from the solar wind and solar energetic particles in lunar ilmenites and pyroxenes. *J. Geophys. Res.* **98**, 13147–13162 (1993).
- Honda, M., McDougall, I., Patterson, D. B., Dougeris, A. & Clague, D. A. Possible solar noble gas component in Hawaiian basalts. *Nature* **349**, 149–151 (1991).
- Honda, M., McDougall, I., Patterson, D. B., Dougeris, A. & Clague, D. A. Noble gases in submarine pillow basalt glasses from Loihi and Kilauea, Hawaii: a solar component in the Earth. *Geochim. Cosmochim. Acta* **57**, 859–874 (1993).
- Hiyagon, H., Ozima, M., Marty, B., Zashu, S. & Sakai, H. Noble gases in submarine glasses from mid-oceanic ridges and Loihi seamount: constraint on the early history of the Earth. *Geochim. Cosmochim. Acta* **56**, 1301–1316 (1992).
- Staudacher, R., Sarda, P. & Allègre, C. J. Noble gas systematics of Réunion Island, Indian Ocean. *Chem. Geol.* **89**, 1–17 (1990).
- Valbracht, P. J. et al. Helium, neon and argon isotope systematics in Kerguelen ultramafic xenoliths: implication for mantle source signatures. *Earth Planet. Sci. Lett.* **138**, 29–38 (1996).
- Poreda, R. J. & Farley, K. A. Rare gases in Samoan xenoliths. *Earth Planet. Sci. Lett.* **113**, 129–144 (1992).
- Kennedy, B. M., Lynch, M. A., Reynolds, J. H. & Smith, S. P. Intensive sampling of noble gases in fluids at Yellowstone: I. Early overview of the data; regional pattern. *Geochim. Cosmochim. Acta* **49**, 1251–1261 (1985).
- Wellman, P. & McDougall, I. Cainozoic igneous activity in eastern Australia. *Tectonophysics* **23**, 49–45 (1974).
- McDonough, W. F., McCulloch, M. T. & Sun, S. S. Isotopic and geochemical systematics in Tertiary-Recent basalts from southeastern Australia and implications for the evolution of the sub-continental lithosphere. *Geochim. Cosmochim. Acta* **49**, 2051–2067 (1985).
- O'Reilly, S. Y. & Zhang, M. Geochemical characteristics of lava-field basalts from eastern Australia and inferred sources: connections with the subcontinental lithospheric mantle? *Contrib. Mineral. Petrol.* **121**, 148–170 (1995).
- Cull, J. P. An appraisal of Australian heatflow data. *Bur. Miner. Resour. J. Austr. Geol. Geophys.* **7**, 11–21 (1982).
- Cull, J. P. & Conley, D. Geothermal gradients and heat flow in Australian Sedimentary basins. *Bur. Miner. Resour. J. Austr. Geol. Geophys.* **8**, 329–337 (1983).
- Zielhuis, A. & Hilst, R. D. v. d. Upper mantle shear velocity beneath eastern Australia from inversion of waveforms from SKIPPY portable arrays. *Geophys. J. Int.* **127**, 1–16 (1996).
- Wass, S. Y. in *The Mantle Sample: Inclusions in Kimberlites and Other Volcanics* (Proc. 2nd Int. Kimberlite Conf., 2) (eds Boyd, F. R. & Meyer, H. O. A.) 366–373 (Am. Geophys. Un., Washington, 1979).
- Anderson, T., O'Reilly, S. Y. & Griffin, W. L. The trapped fluid phase from Victoria, Australia: implications for mantle metasomatism. *Contrib. Mineral. Petrol.* **88**, 72–85 (1984).
- O'Reilly, S. Y. in *Mantle Xenoliths* (ed. Nixon, P. H.) 661–670 (Wiley, Chichester, 1987).
- Yaxley, G. M., Crawford, A. J. & Green, D. H. Evidence for carbonate metasomatism in spinel peridotite xenoliths from western Victoria, Australia. *Earth Planet. Sci. Lett.* **107**, 305–317 (1991).
- Griffin, W. L., O'Reilly, S. Y. & Stabel, A. Mantle metasomatism beneath western Victoria, Australia: II. Isotopic geochemistry of Cr-diopside lherzolites and Al-augite pyroxenites. *Geochim. Cosmochim. Acta* **52**, 449–459 (1988).
- Wetherill, G. W. Variations in the isotopic abundances of neon and argon extracted from radioactive minerals. *Phys. Rev.* **96**, 679–683 (1954).
- Hünemohr, H. Edelgase in U- und Th-reichen Mineralen und die Bestimmung der ²¹Ne-Dicktarget-Ausbeute der ¹⁸O(α,n)²¹Ne-kernreaktion im Bereich 4.0–8.8 MeV. Thesis, Johannes-Gutenberg-Universität (1989).
- Yatsevich, I. & Honda, M. Production of nucleogenic neon in the Earth from natural radioactive decay. *J. Geophys. Res.* **102**, 10291–10298 (1997).
- Barton, C. E. & McElhinny, M. W. A 10000 yr geomagnetic secular variation record from three Australian maars. *Geophys. J. R. Astron. Soc.* **67**, 465–485 (1981).

Acknowledgements. We thank W. Griffin for advising on appropriate samples for this study, and for comments on the manuscript; J. M. McCarron for providing the Mr Gambier samples; and G. Yaxley and M. Handler for comments.

Correspondence should be addressed to T.M. (e-mail: matsumoto@ess.sci.osaka-u.ac.jp).

How aggressive ant-guards assist seed-set in *Acacia* flowers

P. G. Willmer & G. N. Stone*

School of Biological & Medical Science, University of St Andrews, Fife KY16 9TS, UK

* Department of Zoology, South Parks Road, Oxford OX1 3PS, UK

The phenomenon of ant-guarding on *Acacia* trees is probably the best known case of a mutualism between plants and animals, the ants conferring biotic defence against herbivores and perhaps against encroaching vegetation^{1–3}. However, as with many defence mutualisms, sometimes the interests of the plant and its defender conflict: for example, when they are in flower the *Acacia* trees require the presence and service of other insects to effect cross-pollination. How is pollinator access achieved in the face of aggressive ant-guards? Here we report that ants are deterred from young flowers at the crucial stage of dehiscence, allowing bees and other pollinators to visit and transfer pollen. This deterrence appears to be a response to a volatile chemical signal from young flowers, perhaps from the pollen itself. Ants patrol the young (undeveloped) buds, and also return to the flowers after dehiscence, protecting the fertilized ovules and developing seeds. The outcome is a directly improved seed-set in the presence of ants (rather than an indirect extra reproductive resource allocation due to decreased defoliation^{4–6}).

Acacia zanzibarica and *A. drepanolobium* are classic ant-guarded trees of eastern Africa⁷; ants (such as *Crematogaster*⁸) reside on the trees inside modified thorns (pseudogalls) and patrol the branches, attacking any insect or vertebrate herbivores encountered. As in other guarded *Acacias*^{1–3}, herbivory is thereby much reduced. Cross-pollination could be allowed through reduced ant activity during flowering but, as flowers appear soon after seasonal rains and accompany the flush of vegetative growth, this would leave the tree foliage and the flowers themselves vulnerable to damage.

A. zanzibarica and *A. drepanolobium* flowers occur in spherical inflorescences of 20–100 florets bearing conspicuous exposed anthers. Flowering occurs mainly in November–January in Mkomazi. The flowers open at around 05:00 and start to dehisce at 08:00–09:00 (ref. 9). They appear pristine through day 1, persisting through day 2 in a tattered non-functional state before being shed, leaving a small proportion of ovules effectively fertilized. Seed-set is apparent within 2–3 days, although seeds take about a month to mature.

Flower visitors operate on clear diurnal patterns⁹. As in other *Acacias*¹⁰, most pollen collection is carried out by solitary bees (Megachilidae and Halictidae), making 98% of visits to *A. drepanolobium* and 40% of visits to *A. zanzibarica*. Bee activity begins from first dehiscence and peaks at 11:00–14:00 (ref. 9); around 75–80% of available pollen is removed within these 3 hours. The flowers contain little or no nectar and, unlike some sympatric *Acacia* species, are therefore relatively unattractive to nectarivorous pollinators such as butterflies⁹. However, flower predation by beetles (meloids and scarabaeids) is a significant problem, with anthers, stigmas and ovaries being chewed in about 10% of flowers.

Ants also show a daily activity pattern (Table 1), patrolling twigs adjacent to their pseudogall in the early morning, declining in numbers from 09:00–10:00, only rarely in evidence until around 14:00, then increasing during most of the afternoon. Therefore, there are strictly defined times when ants are not present on flowers, corresponding to peak dehiscence and visitation of flowers. Because of the temporal patterning of their behaviours, pollinators meet

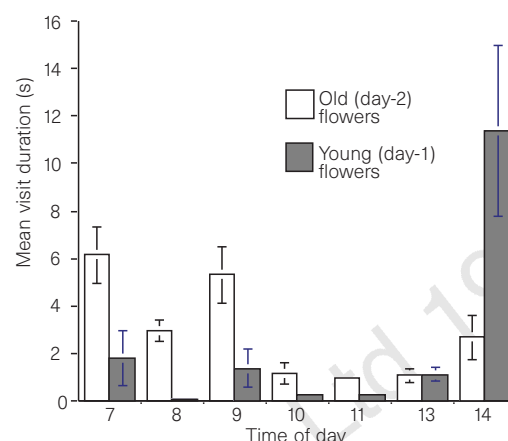


Figure 1 Duration of visits of young and old flowers of *Acacia zanzibarica* through the diurnal cycle, as mean values ($\pm$ s.e.m.) from 4 days of observation.

Table 1 Numbers of patrolling ants

Time	<i>A. zanzibarica</i>			<i>A. drepanolobium</i>
	22 Nov	24 Nov	17 Dec*	21 Dec
06:00–07:00	–	–	9.5	–
07:00–08:00	6.7	13.7	7.2	19.2
08:00–09:00	3.8	8.6	6.9	11.0
09:00–10:00	1.8	4.7	5.9	5.4
10:00–11:00	0.4	3.3	2.3	4.0
11:00–12:00	1.1	3.2	2.4	4.4
12:00–13:00	0.9	1.5	2.4	1.7
13:00–14:00	0.9	2.3	0.3	1.1
13:00–15:00	2.7	3.7	0.4	4.2
15:00–16:00	3.4	7.5	0.7	5.0
16:00–17:00	3.7	–	2.6	5.7
17:00–18:00	3.1	10.9	6.9	7.1

The mean numbers of ants patrolling per 10 marked 0.5-m lengths of branch of different *Acacia* trees on different days in 1995.

* Trees not in flower.

Table 2 ANOVA for duration of ant visits

Source	DF	Seq SS	Adj MS	F	P
Age	1	43,917	1,999	0.20	0.652
Time	8	229,084	23,627	4.83	0.000
Age $\times$ time	8	128,042	16,005	3.27	0.001
Error	265	1,295,876	4,890		
Total	282	1,696,919			

The analysis of variance is shown for the duration of ant visits with age of flower and time. Seq SS, Sequential sum of squares; Adj MS, adjusted mean squares. DF, degrees of freedom.

relatively few ants on the flowers. Nevertheless, ants and pollinators such as bees do overlap in their flower-visiting activities, especially at the onset of dehiscence. How then is conflict avoided?

Ant activities in relation to flower age (Fig. 1) show that ant–flower interactions are more subtle than indicated by an assessment of mere presence or absence. In the early morning, the ants visit more of, and make longer visits to, older (day-2) flowers and the unopened buds (not shown), but significantly avoid younger flowers and stay only briefly (often for 1–5 seconds) when they do encounter a young flower. During such visits, they also differ in their behaviour, with their antennae and mandibles being unusually active and the abdomen often being cocked upwards. Ants could be reacting to flower occupancy by pollinators, but ‘occupied’ flowers are rare at these times (07:00–10:00) and ants make equally short visits to both occupied (6.47 ± 7.20 s; $n = 11$) and unoccupied (7.15 ± 5.45 s; $n = 75$) young flowers. In contrast, when ant activity increases again in the afternoon, younger flowers are intensively

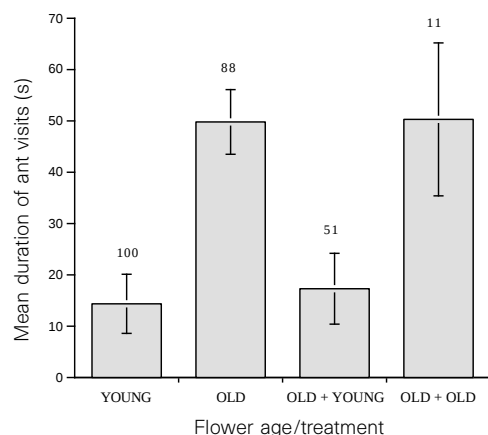


Figure 2 Duration of visits (mean $\pm$ s.e.m.; n above bars) between 07:00–09:30 from two days of observation with four age/treatment groups of *A. zanzibarica* flowers (see Methods).

visited and for significantly longer than during morning visits (Table 2: age, not significant; time, $P = 0.000$; time/age interaction, $P = 0.005$). These patterns could be explained by young flowers producing a deterrent to ants when they first dehisce.

The changing response of ants to young flowers suggests that there is a progressive decay of ant-repellant stimulus after dehiscence. This could result from a short-lived volatile chemical signal from the young flowers. Some flowers do contain ant-repellant nectar^{11,12}, but this is an implausible source of repellent in *Acacia zanzibarica*, where nectar volumes are negligible. The repellent must be present either on the corolla surfaces or on the pollen itself. It might therefore be transferable by contact between flowers, either by direct wiping of chemical(s) or by transfer of whole pollen grains/polyads. To test this, we watched ant behaviour on selected *A. zanzibarica* trees, treated as described under Methods. Figure 2 shows that old flowers receive longer morning visits than young ones, as expected (Fig. 1); wiping old flowers with other old flowers has no effect, whereas wiping with young flowers produces avoidance and shorter visits, as for untreated young flowers (ANOVA, $F = 17.63$, $P = 0.000$), and also elicits an exaggerated behavioural response. Some component transferable from the young dehiscing flowers (therefore a chemical rather than a tactile cue) is acting as an ant deterrent. If freshly dehiscing flowers have a volatile ant-deterrent, this also explains why just one day of observation (27 November 1995) gave ant-visit durations that were invariant between young (40 ± 11 ; $n = 12$) and old (34 ± 6 ; $n = 9$) flowers; rain on that morning had presumably washed away the deterrent cue.

In this model, ants could aid *Acacia* pollination by keeping the very young flower buds pristine and uneaten, then by their absence allowing access to pollinators at peak flower fertility (with the small cost of losing some young flowers), and finally by returning to deter flower-feeding beetles and seed predators. Therefore, if ant abundance per branch is consistent over the flowering and fruiting

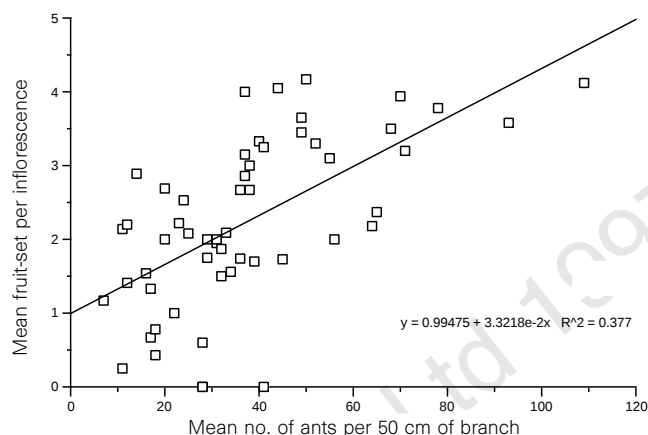


Figure 3 Effects of ant-guard abundance on seed-set. Ant abundance is the number per branch for 5 branches on each of 11 trees; seed-set was scored on the same branch lengths. The line shown is the best-fit linear regression.

period, it should be positively related to eventual seed-set on a given branch. Counts of ants per branch on 10 branches from a single tree over 4 days show a highly significant rank effect of branch (Friedman, $P = 0.006$), indicating that each branch has a rather fixed number of patrollers over time (as expected, given the presence of pseudogalls of fixed size and number). To show the effects of ant activities on overall pollination and seed-set, we therefore compared five branches from each of eleven *A. zanzibarica* trees (Fig. 3). The mean ant abundance per branch is positively correlated with the mean fruit-set per inflorescence for each of the 11 trees. Using a repeated measure ANCOVA (Table 3) with 'trees' as units and 'branches' as repeats, the relationship between ant activity and fruit-set is highly significant ($P = 0.000$). None of the other 'non-ant' factors measured (see Methods) is a significant predictor of fruit-set. Nor are there likely to be environmental factors differing between the 11 adjacent trees that could independently affect both ant abundance and fruit-set^{5,6}. Therefore the presence of guarding ants *per se* can be taken to assist fruit-set in this *Acacia* species. □

Methods

All studies were carried out on *Acacia* trees in the western end of Mkomazi Game Reserve, Northern Tanzania. Flower visiting was scored by quiet continuous observation of tagged 0.6-m branch lengths, with individual flowers noted and aged. To assess the consistency of ant activity patterns, patrolling ants were scored at different times of day over 4 days (on 22, 24, 27 and 28 November 1995) on ten 0.5-m tagged lengths of branch on one tree.

At about 07:00 on two days, we set up and tagged flowers, treated in the following categories: untouched young flowers; untouched old flowers; young flowers 'wiped' onto old flowers; old flowers 'wiped' onto old flowers, as a control for handling effects. Wiping involved smearing the recipient flower's surfaces with either a young or an old flower from the same tree; the picked 'donor' flower, held in forceps, was moved three times over the recipient flower to give contact with most of the spherical exposed surface of massed anthers and stigmas.

To assess ant effects on fruit-set, we selected eleven *A. zanzibarica* trees, all growing close together in a small stand on flat ground to reduce edaphic variations, and tagged equal 0.5-m lengths of five branches on each tree. We then scored ant abundance on each branch totalled across 12 sampling times (07:30–14:30) on 2 December 1995. Numbers and sizes of pseudogalls on that branch were also scored, together with numbers of flowers of each age, the height, aspect and shading of branch, and the tree size (height and trunk girth), to rule out other possible causes of variation in seed-set. Finally, we counted all seeds on each former flower-head site separately (range, 0–11 per flower site, with 0–39 sites per branch).

Table 3 Repeated-measures ANCOVA for fruitset

Source	DF	Seq SS	Adj MS	F	P
Tree	10	657.5	31.5	7.58	0.000
Ants	1	79.4	79.4	19.08	0.000
Error	40	166.4	4.2		
Total	51	903.3			

The tree is subject factor and the branch is within-subject factor, weighted for the numbers of observations. Abbreviations as Table 2.

Received 20 March; accepted 24 April 1997.

- Janzen, D. H. Coevolution of mutualism between ants and *Acacia* in Central America. *Evolution* **20**, 249–275 (1966).
- Hocking, B. Insect associations with the swollen-thorn acacias. *Trans. R. Ent. Soc. Lond.* **122**, 211–255.
- Knox, R. B., Marginson, R., Kenrick, J. & Beattie, A. J. in *Insects and the Plant Surface* (eds Juniper, B. E. & Southwood, T. R. E.) 295–307 (Arnold, London, 1986).
- Del-Claro, K., Berto, V. & Reu, W. Effect of herbivore deterrence by ants on the fruit set of an extrafloral nectary plant *Qualea multiflora* (Vochysiaceae). *J. Trop. Ecol.* **12**, 887–792 (1996).
- Niesenbaum, R. A. Linking herbivory and pollination: defoliation and selective fruit abortion in *Lindera benzoin*. *Ecology* **77**, 2324–2331 (1996).
- Koptur, S. in *Insect-Plant Interactions* (ed. Bernays, E.) 81–129 (CRC Press, Boca Raton, 1992).
- Ross, J. H. An analysis of the African *Acacia* species: their distribution, possible origins and relationships. *Bothalia* **13**, 389–413 (1981).
- Young, T. P., Stubblefield, C. H. & Isbell, L. A. Ants on swollen thorn *Acacias*: species coexistence in a simple system. *Oecologia* **109**, 98–107 (1997).
- Stone, G. N., Willmer, P. G. & Nee, S. Daily partitioning of pollinators in an African *Acacia* community. *Proc. R. Soc. Lond. B* **263**, 1389–1393 (1996).
- Tyrbirk, K. Pollination, breeding system and seed abortion in some African acacias. *Bot. J. Linn. Soc.* **112**, 107–137 (1993).
- Feinsinger, P. & Swann, L. A. How common are ant-repellent nectars? *Biotropica* **10**, 238–139 (1978).
- Prys-Jones, O. E. & Willmer, P. G. The biology of alkaline nectar in the Purple Toothwort (*Lathraea clandestina*): ground level defence. *Biol. J. Linn. Soc.* **45**, 373–388 (1992).

Acknowledgements. We thank D. Mafunde at Mkomazi Research Station for expert field assistance; T. Morgan and N. McWilliam for invaluable logistic back-up; and M. Ritchie and J. Graves for statistical advice. We thank the Royal Geographical Society, the Darwin Initiative and the British Council for financial support, as part of the Mkomazi Ecological Research Programme, and the Department of Wildlife of the Tanzanian Government.

Correspondence should be addressed to P.G.W. (e-mail: pgw@st-andrews.ac.uk).

Evolution of genetic redundancy

Martin A. Nowak*, Maarten C. Boerlijst*, Jonathan Cooke† & John Maynard Smith‡

* Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

† National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK

‡ School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

Genetic redundancy means that two or more genes are performing the same function and that inactivation of one of these genes has little or no effect on the biological phenotype. Redundancy seems to be widespread in genomes of higher organisms^{1–9}. Examples of apparently redundant genes come from numerous studies of developmental biology^{10–15}, immunology^{16,17}, neurobiology^{18,19} and the cell cycle^{20,21}. Yet there is a problem: genes encoding functional proteins must be under selection pressure. If a gene was truly redundant then it would not be protected against the accumulation of deleterious mutations. A widespread view is therefore that such redundancy cannot be evolutionarily stable. Here we develop a simple genetic model to analyse selection pressures acting on redundant genes. We present four cases that can explain why genetic redundancy is common. In three cases, redundancy is even evolutionarily stable. Our theory provides a framework for exploring the evolution of genetic organization.

There are an increasing number of observations demonstrating that experimental inactivation of certain genes has no apparent effect on the phenotype or fitness of an animal. In specific cases, it seems that the natural function of a gene can be taken over by another gene. Such a redundant genetic organization is sensible from an engineer's point of view: important functions require back-up devices that can take over in case of failure. But can natural selection favour the emergence and stability of redundant genes?

Consider a population of animals in which some essential function can be performed by genes at either of two loci, *A* and *B*. (We use the word 'function' to refer to an effect of a gene during development; thus two genes coding for different proteins can have the same function.) Non-functional alleles, *a* and *b*, arise by mutation at rates u_a and u_b per generation; reverse mutations are ignored. For simplicity, we consider a haploid population, but the

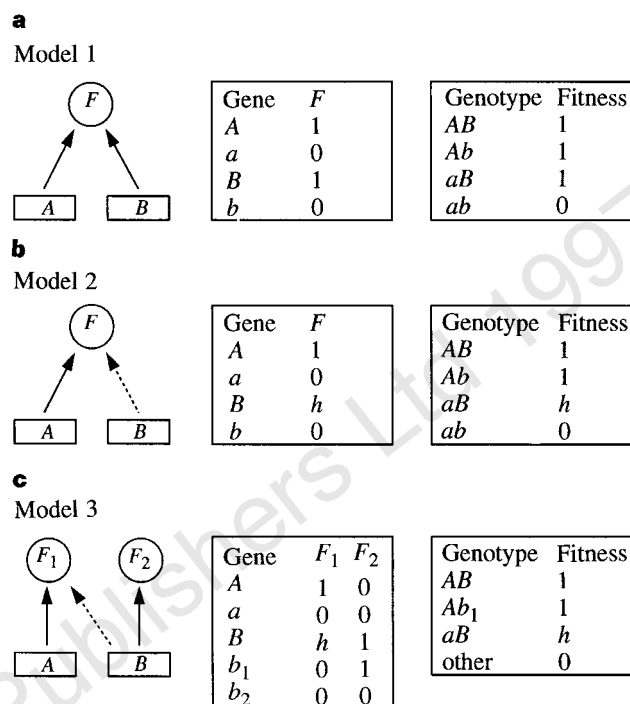


Figure 1 Models of genetic redundancy. **a**, Model 1. If genes *A* and *B* perform function *F* with equal efficacy, then redundancy does not persist. For unequal mutation rates, the gene with higher mutation rate will become extinct. **b**, Model 2. Redundancy can be evolutionarily stable if *B* performs the function with lower efficacy, but also has a lower mutation rate. **c**, Model 3. Suppose *A* and *B* perform different functions, *F*₁ and *F*₂, but *B* also performs function *F*₁, but with lower efficacy than gene *A*. The redundant organization for performing function *F*₁ is evolutionarily stable if the mutation rate that eliminates the cross-function in *B* (but leaves its primary function unaffected) is lower than the mutation rate that inactivates *A*.

models can be extended to diploid populations and the conclusions remain essentially unchanged. There are four genotypes: *AB*, *Ab*, *aB* and *ab*. In each generation, random mating is followed by mutation and selection. Natural selection can maintain both genes if redundancy is only apparent, that is, if the *AB* genotype is fitter than the other genotypes. Less obvious is the question of whether natural selection can maintain true redundancy in the sense that an individual with one of the two genes is as fit as an individual with both. Models 1–3 will address this question. Model 4 studies the consequence of developmental errors.

In model 1, we assume that both genes are equally effective, and that each can function perfectly on its own (Fig. 1a). The fitness of *AB*, *Ab* and *aB* is one, while the fitness of *ab* is zero. Let us first consider the case where the mutation rates in both genes are the same: $u_a = u_b = u$. The system admits a line of equilibria. All trajectories converge to this line. For small mutation rates, the maximum equilibrium frequency of *AB* is approximately $1 - 2/(u/r)$, where r is the recombination rate between the two loci. Thus a large proportion of individuals can carry functional alleles for both genes.

There is, however, an important caveat. We have assumed that the mutation rates u_a and u_b are equal. But any small deviation from $u_a = u_b$ destroys the equilibrium line. If $u_a \neq u_b$ then model 1 does not admit any interior equilibrium, and redundancy does not survive²². A simple way of understanding this result is as follows. At equilibrium, the rate at which deleterious genes arise by mutation must equal the rate at which they are removed by selection. Because

only ab individuals are removed selectively, the rates of removal of the two genes are equal: therefore the rates at which they arise by mutation must be equal.

If $u_a > u_b$, then A will become extinct while B will be fixed. But if mutation rates are very similar, this may take a long time. In model 1, A declines as $\exp[-(u_a - u_b)T]$, where T is the number of generations²². (This represents an upper limit.) For a mutation rate of 10^{-6} per gene per generation, and a 10% difference between u_a and u_b , the average lifetime of redundancy is about 10^7 generations. Therefore, a certain amount of redundancy in our genomes could be the consequence of recent gene duplication events. We note, however, that redundancy cannot only be a consequence of gene duplication because very different genes can also show overlapping redundancy.

Several authors have studied stochastic versions of model 1 and computed the time it takes for random drift to eliminate one of the two genes even if mutation rates are exactly equal^{23–28}.

In model 2, we assume that the genes A and B perform the same function, but with slightly different efficacies (Fig. 1b). Suppose A performs the function with an efficacy of one, while B does it with a

reduced efficacy, h . If both genes are present, the function is performed with the higher of the two efficacies; this is essentially a definition of redundancy. Thus the fitness of genotypes AB and Ab is one, while the fitness of genotype aB is h . The fitness of ab is zero. Unexpectedly, this can lead to a stable equilibrium with both genes A and B maintained in the population, provided that the mutation rate in A is higher than in B . Redundancy is maintained because gene B , with the lower efficacy, also has a lower mutation rate, and is maintained by selection in a genotypes. In this case, B is fully redundant in the sense that its inactivation has no effect on fitness, whereas deletion of A causes a small reduction in fitness. A stable equilibrium is also possible if the fitness of Ab is higher than the fitness of AB , which in turn has a higher fitness than aB . In this case, redundancy is even maintained at a cost.

Model 3 relates pleiotropy to redundancy (Fig. 1c). Pleiotropy implies that genes perform more than one specific function (for example, by being expressed at more than one time and place in the developing organism²⁹). The idea is that redundancy between two genes occurs only with respect to a given function, while the genes are maintained by selection because of another, independent func-

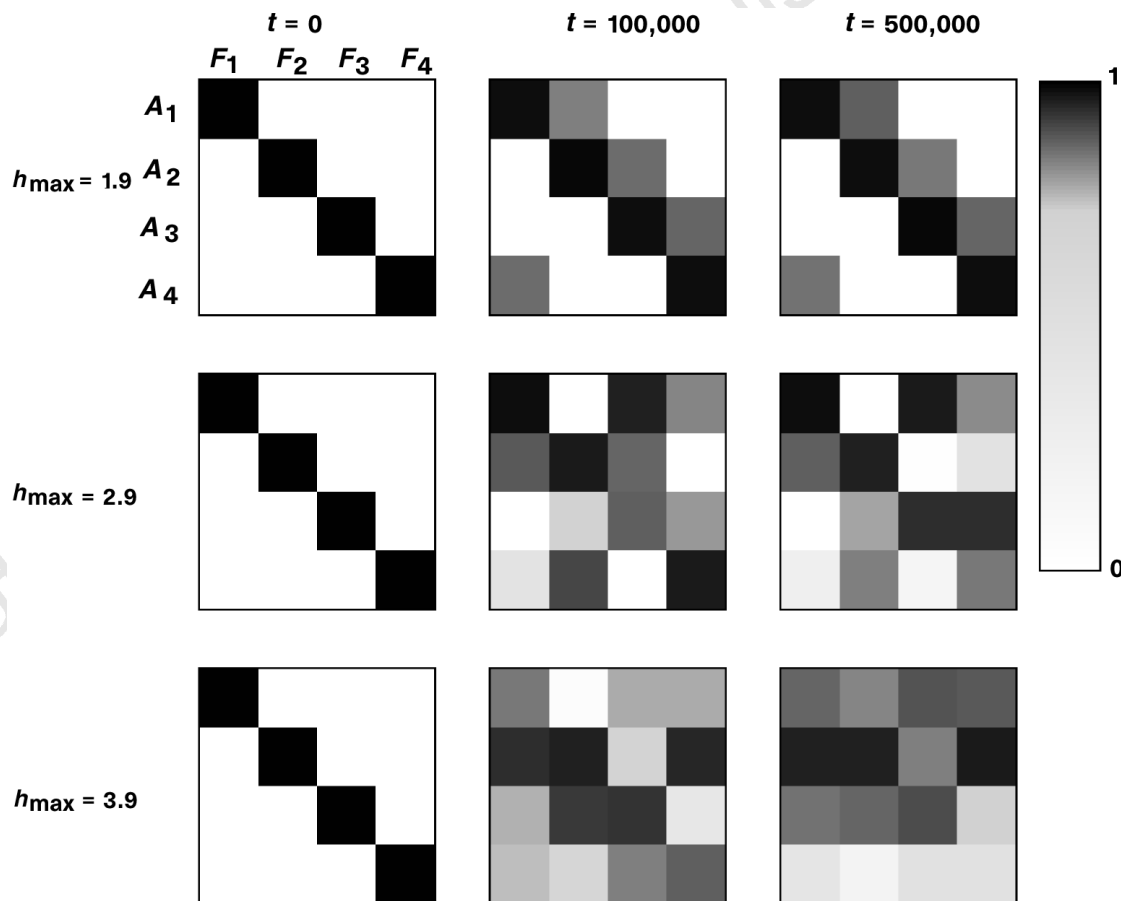


Figure 2 Complex redundancy-pleiotropy networks evolve if the mutation rate of complete inactivation of a gene is higher than the mutation rate of inactivating only one function of a gene while leaving other functions unaffected. Results from a stochastic computer simulation are shown. In each round, two individuals are chosen to reproduce, and their genes are recombined and mutated. Mutation can either inactivate a gene completely, inactivate one or more functions of a gene, or change the efficacies at which a gene performs a function. The fitness of the offspring is given by the product of the efficacies at which each function is performed, where each function is performed with the efficacy of the most efficient gene for this function. There are four loci, A_1 to A_4 , and four functions, F_1 to F_4 . Gene i performs function j with efficacy h_{ij} . Initially each gene performs one

function with an efficacy of one. During the simulation, each gene can evolve to perform additional functions, but the sum over all efficacies is limited by $h_{\max}$ for each gene: $\sum_j h_{ij} < h_{\max}$. For example, $h_{\max} = 1.9$ means that a gene can perform one function with an efficacy of one and a second function with an efficacy of 0.9. The figure shows the initial configuration, and the population averages of h_{ij} after $t = 100,000$ and $t = 500,000$ generations for three different values of $h_{\max}$. The population size is 20,000. The mutation rate for complete inactivation is $u_1 = 0.001$, $u_2 = 0.0011$, $u_3 = 0.0012$ and $u_4 = 0.0013$ for loci 1–4. A specific function is lost with mutation rate $u_i/10$ and changed to random value between 0 and 1 (subject to $\sum_j h_{ij} < h_{\max}$) with mutation rate $u_i/20$.

tion. Redundancy arises as a consequence of 'functional overlap' between genes.

In the simplest case, there are two functions, F_1 and F_2 , and two genes, A and B . Suppose A performs F_1 with an efficacy of one, while B performs F_1 with a slightly reduced efficacy, h , and F_2 with an efficacy of one. Mutations in A lead to the inactive variant a ; the mutation rate is u_a . In the second locus, we consider two types of mutants: b_1 has lost the ability to perform F_1 , but still performs F_2 ; b_2 is completely inactive. The mutation rate from B to b_1 is u_{b1} .

The fitness of each variant is evaluated by assuming that each function is performed with the efficacy of the most efficient gene, and the overall fitness is the product of the efficacies at which the two functions are performed. Therefore genotypes AB and Ab_1 have a fitness of one, genotype aB has fitness h , and all other genotypes have a fitness of zero.

Using the same framework as above, we find that a stable equilibrium with AB is possible provided the mutation rate u_{b1} is smaller than u_a . Functional overlap (and therefore redundancy in performing function F_1) is stable if the mutation rate at which mutants are produced that have lost the functional overlap but still

maintain the original function is lower than the mutation rate of producing inactive mutants at the other locus ($u_{b1} < u_a$). This is plausible, as mutations that destroy all functions of a gene are likely to be more common than mutations destroying one function but leaving another unaffected. This is especially true if the two functions are very similar.

Models 2 and 3 show that true redundancy can be evolutionarily stable, but in both cases the relevant selection pressures are weak if the mutation rates are low. For weak selection pressures to counteract random drift the population size has to be large. In our case the population size has to exceed $1/u$ for redundancy to be maintained.

Models 1–3 explore the maintenance of redundancy. We would also like to understand the origin of redundancy and extend the idea to a larger number of genes and functions. Figure 2 shows computer simulations based on a stochastic version of our model. The starting configuration is neither redundant nor pleiotropic: each function is performed by one gene, and each gene performs one function. During the simulation, mutations that lead to functional overlap arise spontaneously and can be favoured by selection. Genes evolve to perform additional functions. Redundancy is selected and can be fixed in the population if the mutation rate for complete inactivation of a gene is higher than the mutation rate for inactivating only a specific function of a gene. Final configurations often consist of complex redundancy–pleiotropy networks in which each function is performed by several genes and each gene performs several functions.

Finally, model 4 considers 'developmental errors' (Fig. 3). The transmission of information from the egg to the adult organism is subject to errors'. Let us therefore consider the possibility that a gene is intact in the germ line but fails to perform its function during development. We suggest that such developmental failure may arise either by somatic mutation, by errors in the origin and maintenance of cell differentiation (for example, in copying DNA methylation patterns), or through errors in cell-to-cell signalling³⁰.

In terms of our model, we assume that two genes A and B perform the same function, but with probabilities δ_a and δ_b gene A and B , respectively, fail to perform this function in the course of development. If both genes fail, the function is not performed and the animal does not survive.

In the absence of developmental error, the genotypes AB , Ab and aB have a fitness of one, but taking into account developmental error, the average fitnesses of AB , Ab and aB are respectively $1 - \delta_a\delta_b$, $1 - \delta_a$, and $1 - \delta_b$. As before, the mutation rates to produce inactive genes a and b are u_a and u_b . We find that the redundant genotype, AB , is stable if $u_a < \delta_b$ and $u_b < \delta_a$. Thus the mutation rate in each gene has to be smaller than the developmental error rate in the other gene. It is plausible that such developmental failures are more frequent than germline mutations: repeated rounds of cell replication provide increased probabilities for somatic mutation; errors can occur in the DNA methylation pattern and may result in incorrect cell differentiation; and interactions with other cells and response to signals can fail. Therefore we expect the error rate for normal expression of developmental genes, per individual ontogeny, to be higher than germline mutation rates. This is also supported by observations that routine examination of large numbers of embryos in each phase of development reveal spontaneous cases of obvious morphogenetic failure much in excess of germline mutation rates.

Model 4 suggests that redundancy should be more common in developmental genes that are expressed in specific spatio-temporal patterns in the body than in genes encoding for 'housekeeping' functions that are required in all cells (for example, essential metabolic enzymes). Somatic mutations or failures of gene expression that simply kill the cell in which they occur may have much less phenotypic effect than similar events that misguide subsequent developmental signals. Therefore the developmental error rate should be higher in genes that are not required in every cell.

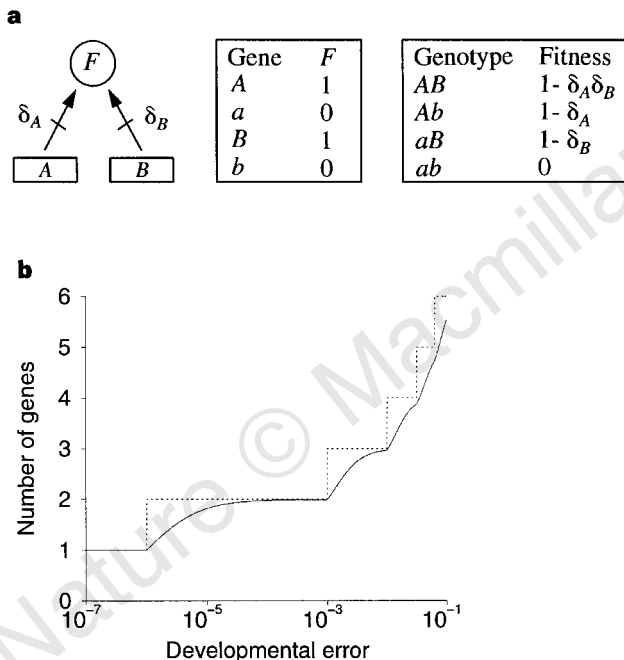


Figure 3 Developmental errors can provide selection pressure to maintain genetic redundancy. **a**, Consider two genes A and B with mutation rates u_a and u_b and developmental error rates δ_a and δ_b . Both genes are maintained in the population provided that $u_a < \delta_b$ and $u_b < \delta_a$. **b**, The model can be extended to n genes. Suppose all genes have mutation rate u and developmental error rate δ . The step function shows the maximum number of redundant genes that can coexist for a given development error rate δ . The continuous function indicates the population average, Σx_i , where x_i is the frequency of genotypes with i functional genes. The mutation rate is $u = 10^{-6}$.

The model can be extended to more than two genes per function (Fig. 3b). An elegant result is obtained if we consider genes with similar mutation rate, u , and similar developmental error rate, δ . The number of redundant genes that can be maintained per function is the largest integer less than $1 + (\log u)/(\log \delta)$. For example, if the mutation rate is $u = 10^{-6}$ and the developmental error rate is $\delta = 10^{-3}$, selection can maintain up to three genes for a given function.

Model 1 shows that in situations where true redundancy is not evolutionarily stable, it may nevertheless take a long time until it is eliminated from the population, provided that mutation rates are small. Models 2 and 3 describe situations in which true redundancy can be maintained indefinitely. Model 3 can lead to complex redundancy networks in which each function is performed by several genes and each gene performs more than one function. Such networks are evolutionarily stable provided that random mutations are more likely to destroy all functions of a gene, rather than destroy just one function while leaving other functions unaffected. Model 4 introduces the concept of developmental errors, and shows that redundancy is evolutionarily stable provided that developmental error rates are larger than mutation rates in the germ line. According to model 4, redundancy should occur for those genes (or functions) that are under a high developmental error rate. The four models are not mutually exclusive; together they explain how mutation and selection can lead to redundant genetic organization. □

Methods

Model 1. Consider a haploid population with genes at two loci, A and B . Non-functional alleles, a and b , arise at mutation rates u_a and u_b . There are four genotypes, AB , Ab , aB and ab . The frequencies are x_1, x_2, x_3 and x_4 , and the fitnesses are f_1, f_2, f_3 and f_4 , respectively. In each generation there is mating (with recombination), followed by mutation and selection. Mating is described by the difference equations: $x'_1 = x_1 + D$, $x'_2 = x_2 - D$, $x'_3 = x_3 - D$, and $x'_4 = x_4 + D$. Here, $D = r(x_2x_3 - x_1x_4)$, where r is the recombination rate between the A and B loci, and r is a number between 0 and 0.5. Mutation is described by $x'_1 = x_1(1 - u_a)(1 - u_b)$, $x'_2 = x_1(1 - u_a)u_b + x_2(1 - u_a)$, $x'_3 = x_1u_a(1 - u_b) + x_3(1 - u_b)$, and $x'_4 = x_1u_au_b + x_2u_a + x_3u_b + x_4$. Selection is described by $x'_i = f_i x_i / \bar{f}$, where $\bar{f} = \sum x_i f_i$ denotes the average fitness of the population. Suppose both genes perform function F with equal efficacy. We have $f_1 = f_2 = f_3 = 1$ and $f_4 = 0$. For exactly equal mutation rates, $u_a = u_b = u$, there is a line of equilibria given by $x_1 = x_2x_3r(1 - u)/u$. For unequal mutation rates, the gene with higher mutation rate will become extinct.

Model 2. This has the same framework as model 1, but genes A and B perform function F with different efficacies, h_a and h_b . Let $h_a > h_b$. The genotype fitnesses are $f_1 = f_2 = h_a$, $f_3 = h_b$ and $f_4 = 0$. Redundancy can be evolutionarily stable if B has a lower mutation rate than A , $u_b < u_a$. If $1 - (h_b/h_a) > u_a > u_b[1 + (1/r)(h_a - h_b)/h_b]$ the equilibrium is $x'_1 = (1 - x'_2) \times [h_a(1 - u_a) - h_b(1 - u_b)] / [(h_a - h_b)(1 - u_a)]$, $x'_2 = (1/r)[u_b/(1 - u_b) \times [h_a(1 - u_a) - h_b(1 - u_b)] / [h_b(u_a - u_b)]]$, $x'_3 = 1 - x'_1 - x'_2$, and $x'_4 = 0$. For low mutation rates, the equilibrium frequency of the redundant AB genotype is approximately $x'_1 \approx 1 - (1/r)[u_b/(u_a - u_b)](h_a - h_b)/h_b$. For example, if $h_a = 1$, $h_b = 0.99$, $u_a = 1.1 \times 10^{-6}$, $u_b = 10^{-6}$ and $r = 0.5$, then the equilibrium frequency of AB is about 0.8.

This model can be expanded to n genes with different mutation rates and different efficacies. The fitness of a particular genotype is given by the efficacy of the most efficient gene. If less efficient genes have lower mutation rates then stability of several redundant genes is possible. For a large number of genes, however, the conditions on efficacies and mutation rates become very restrictive.

Model 3. Consider two genes, A and B , and two functions, F_1 and F_2 . Gene A performs function F_1 with efficacy h_a , and gene B performs function F_1 with a lower efficacy h_b and function F_2 with an efficacy of one. Mutations in A lead to the inactive variant a ; the mutation rate is u_a . Mutations in B can either lead to variant b_1 , which has lost the ability to perform function F_1 but still performs F_2 , or to variant b_2 , which is completely inactive; mutation rates are u_{b1} and u_{b2} ,

respectively. Variant b_2 can also arise from b_1 at a mutation rate u_{b3} . The redundant organization for performing function F_1 , is evolutionarily stable if $u_{b1} < u_a$. The analysis is similar to model 2 if $u_{b2} \approx u_{b3}$: for low mutation rates, the equilibrium frequency of AB is approximately $x'_1 \approx 1 - (1/r)[u_{b1}/(u_a - u_{b1})] \times (h_a - h_b)/h_a$. For the same numerical values as model 2, and assuming that u_{b1} is 10 times smaller than u_a , we find that the equilibrium frequency of AB is 0.998. Pleiotropy facilitates redundancy.

Model 4. Consider two genes A and B with mutation rates u_a and u_b and developmental error rates δ_a and δ_b . Mutation and selection are described by the difference equations $x'_1 = (1 - \delta_a\delta_b)(1 - u_a)(1 - u_b)x_1/f$, $x'_2 = (1 - \delta_a) \times (x_1u_b + x_2)/f$, $x'_3 = (1 - \delta_b)(x_1u_a + x_3)/f$, $x'_4 = 0$, where f is such that $x'_1 + x'_2 + x'_3 = 1$. In contrast to models 1–3, recombination is not essential here. The equilibrium frequency of AB is $x_1 = 1/[1 + [u_a(1 - \delta_b)/[\delta_b(1 - \delta_a) - u_a(1 - \delta_a\delta_b)] + [u_b(1 - \delta_a)/[\delta_a(1 - \delta_b) - u_b(1 - \delta_a\delta_b)]]]$. For small values of u and δ , we obtain $x_1 \approx 1/[1 + [u_a/(\delta_b - u_a)] + [u_b/(\delta_a - u_b)]]$. Thus necessary conditions for a large x_1 are $u_a < \delta_b$ and $u_b < \delta_a$.

The model can be extended to n genes. Suppose all genes have mutation rate u and developmental error rate δ . Let x_i denote genotypes with i genes ($i = 0, \dots, n$). The population dynamics are $x'_{n-k} = (f_{n-k}/f) \sum_{i=0}^k x_{n-k+i} \times u^{k-i} x_{n-i}$, where $f_j = (1 - \delta)^j(1 - u)^j$ and f is such that all frequencies add to one. The equilibrium can be solved recursively. An equilibrium with the genotype containing all n redundant genes is possible if $f_n > f_{n-1}$. This leads to $n < 1 + (\log u)/(\log \delta)$.

Diploid models. Our results for haploid models also apply to diploid models. In diploid models, we distinguish four gametes, AB , Ab , aB and ab , which form nine zygotes: AB/AB , AB/Ab , AB/aB , AB/ab , Ab/AB , Ab/Ab , aB/AB , aB/Ab , ab/AB , ab/Ab , ab/aB and ab/ab . For each generation we assume that mutation acts on gamete frequency, then zygotes are formed, selection acts on zygotes, and finally new gametes are formed, including the possibility of recombination. In agreement with haploid model 1, we find that the case where all zygotes have high fitness except ab/ab which has low fitness, does not lead to stable redundancy. Cases similar to models 2 and 3 give stable redundancy. Diploid models with developmental errors also give stable redundancy.

There are some additional cases that can lead to redundancy in diploid models. One such case was discovered by Brookfield: it assumes that the double heterozygote, AB/ab , is as fit as the wild type, AB/AB , but Ab/ab , aB/ab and ab/ab have low fitness¹. In addition, stable redundancy is also possible for partial dominance where all homozygotes have high fitness, the double heterozygote has a lower fitness, the single heterozygotes have still lower fitness, and ab/ab has lowest fitness.

Classification of redundancy. It is helpful to distinguish three types of genetic redundancy. (1) True redundancy¹ denotes the situation where an individual with a redundant genotype, AB , is not fitter than one in which one of the redundant genes has been knocked out, Ab . In model 2, B is truly redundant, but A is not. In cases with pleiotropy, 'true redundancy' implies that the fully redundant genotype is not fitter than a genotype where the pleiotropic function of one gene has been eliminated. (2) 'Generic redundancy' is the case when an AB individual is only occasionally fitter than an Ab individual. This can be the consequence of rare developmental errors. Another possibility is that AB is only fitter than Ab in some environments. (3) 'Almost redundancy' means that the redundant genotype AB is always slightly fitter than any genotype where one of the redundant genes has been knocked out. Of course, the fitness difference should be small if the situation is to be regarded as one of redundancy. Several such examples have been discussed previously⁵.

Received 14 January; accepted 25 April 1997.

1. Brookfield, J. F. Y. Genetic redundancy. *Adv. Genet.* **36**, 137–155 (1997).
2. Brookfield, J. F. Y. Can genes be truly redundant? *Curr. Biol.* **2**, 553–554 (1992).
3. Tautz, D. Redundancies, development and the flow of information. *BioEssays* **14**, 263–266 (1992).
4. Goldstein, D. B. & Holsinger, K. E. Maintenance of polygenic variation in spatially structured populations. *Evolution* **46**, 412–429 (1992).
5. Thomas, J. H. Thinking about genetic redundancy. *Trends Genet.* **9**, 395–399 (1993).
6. Dover, G. A. Evolution of genetic redundancy for advanced players. *Curr. Opin. Genet. Dev.* **3**, 902–910 (1993).
7. Pickett, F. B. & Meeks-Wagner, D. R. Seeing double: appreciating genetic redundancy. *Plant Cell* **7**, 1347–1356 (1995).
8. Bird, A. P. Gene number, noise reduction and biological complexity. *Trends Genet.* **11**, 94–100 (1995).
9. O'Brien, S. J. On estimating function gene number in eukaryotes. *Nature New Biol.* **242**, 52–54 (1973).
10. Kastner, P. et al. Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life? *Cell* **83**, 859–869 (1995).
11. Rudnicki, M. A. et al. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* **71**, 383–390 (1992).

12. Saga, Y. *et al.* Mice develop normally without tenascin. *Genes Dev.* **6**, 1821–1831 (1992).
13. Yang, Y. *et al.* Functional redundancy of the muscle-specific transcription factors Myf5 and myogenin. *Nature* **379**, 823–825 (1996).
14. Joyner, A. L. *et al.* Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. *Science* **251**, 1239–1243 (1991).
15. Laney, J. D. & Biggin, M. D. Redundant control of Ultrabithorax by zeste involves functional levels of zeste protein binding at the Ultrabithorax promoter. *Development* **122**, 2303–2311 (1996).
16. Schorle, H. *et al.* Development and function of T cells in mice rendered interleukin-2 deficient by gene targeting. *Nature* **352**, 621–624 (1991).
17. Taniguchi, T. Cytokine signaling through nonreceptor protein tyrosine kinases. *Science* **268**, 251–255 (1995).
18. Steindler, D. A. *et al.* Tenascin Knockout Mice: Barrels, Boundary Molecules, and Glial Scars. *J. Neurosci.* **15**, 1971–1983 (1995).
19. Pekny, M. *et al.* Mice lacking glial fibrillary acidic protein display astrocytes devoid of intermediate filaments but develop and reproduce normally. *EMBO J.* **14**, 1590–1598 (1995).
20. Reed, S. I. G1-specific cyclins in search of an S-phase promoting factor. *Trends Genet.* **7**, 95–99 (1991).
21. Roche, S. *et al.* Requirement for Src family protein tyrosine kinases in G-2 for fibroblast cell division. *Science* **269**, 1567–1569 (1995).
22. Fisher, R. A. The sheltering of lethals. *Am. Nat.* **69**, 446–455 (1935).
23. Christiansen, F. B. & Frydenberg, O. Selection-mutation balance for two nonallelic recessives producing an inferior double homozygote. *Am. J. Hum. Genet.* **29**, 195–207 (1977).
24. Bailey, G. S. *et al.* Gene duplication in tetraploid fish: model for gene silencing at unlinked loci. *Proc. Natl Acad. Sci. USA* **75**, 5575–5579 (1978).
25. Allendorf, F. W. Protein polymorphism and the rate of loss of duplicate gene expression. *Nature* **272**, 76–78 (1978).
26. Kimura, M. & King, J. L. Fixation of a deleterious allele at one of two duplicate loci by mutation pressure and random drift. *Proc. Natl Acad. Sci. USA* **76**, 2858–2861 (1979).
27. Li, W.-H. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fish. *Genetics* **95**, 237–258 (1980).
28. Ohta, T. Time for spreading of compensatory mutations under gene duplication. *Genetics* **123**, 579–584 (1989).
29. Li, X. & Noll, M. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* **367**, 83–87 (1994).
30. Krakauer, D. C. & Pagel, M. Selection by somatic signals. *Phil. Trans. R. Soc. Lond. B* **351**, 647–658 (1996).

Acknowledgements. We thank D. Krakauer and K. Sigmund for discussion. This work was supported by the Wellcome Trust (M.A.N.) and the European Community (M.C.B.).

Correspondence should be addressed to M.A.N. (e-mail: martin.nowak@zoo.ox.ac.uk).

Distinct cortical areas associated with native and second languages

Karl H. S. Kim^{*†}, Norman R. Relkin[†], Kyoung-Min Lee^{*†} & Joy Hirsch^{*†}

Department of Neurology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA*

† Department of Neurology and Neuroscience, Cornell University Medical College, 1300 York Avenue, New York, New York 10021, USA

The ability to acquire and use several languages selectively is a unique and essential human capacity. Here we investigate the fundamental question of how multiple languages are represented in a human brain. We applied functional magnetic resonance imaging (fMRI) to determine the spatial relationship between native and second languages in the human cortex, and show that within the frontal-lobe language-sensitive regions (Broca's area)^{1–3}, second languages acquired in adulthood ('late' bilingual subjects) are spatially separated from native languages. However, when acquired during the early language acquisition stage of development ('early' bilingual subjects), native and second languages tend to be represented in common frontal cortical areas. In both late and early bilingual subjects, the temporal-lobe language-sensitive regions (Wernicke's area)^{1–3} also show effectively little or no separation of activity based on the age of language acquisition. This discovery of language-specific regions in Broca's area advances our understanding of the cortical representation that underlies multiple language functions.

Indirect evidence for topographic specialization within the language-dominant hemispheres of multilingual subjects has been provided by clinical reports of selective impairments in one or more of several languages as a result of surgery involving the left

perisylvian area⁴. Multilingual patients with complex partial seizure disorders of temporal lobe origin have been reported to shift from a primary to a second language together with ictal progression⁵. Different languages have also been selectively disrupted in polyglots by electrical stimulation of discrete regions of the neocortex of the dominant hemisphere^{6,7}. Changes in the topography of background electroencephalogram (EEG) coherence obtained during translation tasks also suggest spatial separation of cortical regions involved in multiple languages⁸. Although these reports are consistent with the existence of spatially separate representations for each language, such functions have not been localized.

Silent, internally expressive linguistic tasks were performed in two languages by subjects who either acquired conversational fluency in their second languages as young adults ('late' bilinguals) or who acquired two languages simultaneously early in their development ('early' bilinguals) (Table 1). As Broca's and Wernicke's areas are known to perform central roles in human language functions^{1–3,9–12}, we have focused our observations on these cortical areas.

The main findings for a typical 'late' bilingual subject (subject (A)) are shown in Fig. 1. The anterior language area is highlighted by the green box and shown expanded in the inset. Red indicates significant activity during the native language task (English), whereas yellow indicates activity associated with the second language task (French). Two distinct but adjacent centres of activation (+) separated by ~7.9 mm were evident within the inferior frontal gyrus, suggesting that two specific regions served each of the two languages. In the posterior language area of the same subject (Fig. 2), the same tasks yielded centroids of activity with a centre-to-centre spacing of 1.1 mm, less than the width of a voxel, suggesting that similar or identical cortical regions served both languages in this posterior area.

For all six late bilingual subjects, distinct areas of activation were observed for the native and second languages in Broca's area (Table 2a and Fig. 3). The separation between centroids of activity ranged from ~4.5 mm to 9.0 mm within one slice, and the number of voxels for each language was similar for each subject. On the other hand, activity in Wernicke's area (Table 2b) showed centre-to-centre distances between the centre-of-mass centroids ranging from 1.1 to 2.8 mm. The mean centroid distance between the anterior

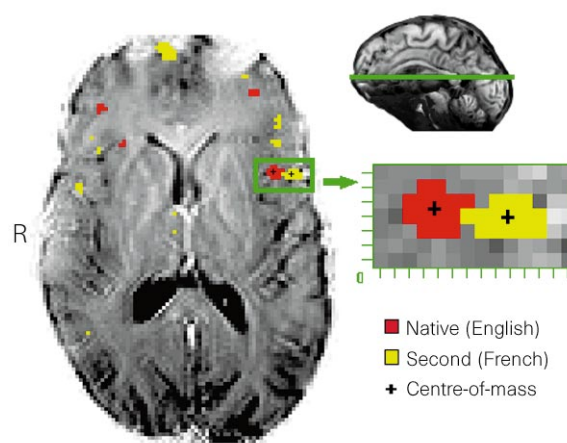


Figure 1 A representative axial slice from a 'late' bilingual subject (A) shows all voxels that pass the multistage statistical criteria at $P \leq 0.0005$ as either red (native language) or yellow (second acquired language). An expanded view of the pattern of activity in the region of interest (inferior frontal gyrus, Brodmann's area 44 (refs 2, 3, 18), corresponding to Broca's area^{1–3}) indicates separate centroids (+) of activity for the two languages. Centre-of-mass calculations indicate that the centroids are separated on this plane by 7.9 mm. The green line on the upper right mid-sagittal view indicates the plane location. R indicates the right side of the brain.

Table 1 Subject information

Subject	Age	Gender	Native language(s)	Second language	Handedness	Laterality quotient ²⁴
A	31	M	English	French	Right	60
B	32	M	Korean	English	Right	100
C	28	M	Korean	English	Right	86
D	26	M	English	Japanese	Ambidextrous	-20
E	27	F	Spanish	English	Right	100
F	32	F	German	English	Right	60
G	38	F	Turkish/English	NA	Ambidextrous	-27
H	27	M	English/Hebrew	NA	Right	85
I	23	M	English/Spanish	NA	Right	100
J	24	F	Croatian/English	NA	Right	89
K	31	M	Italian/German	NA	Right	90
L	32	M	Chinese/English	NA	Ambidextrous	24

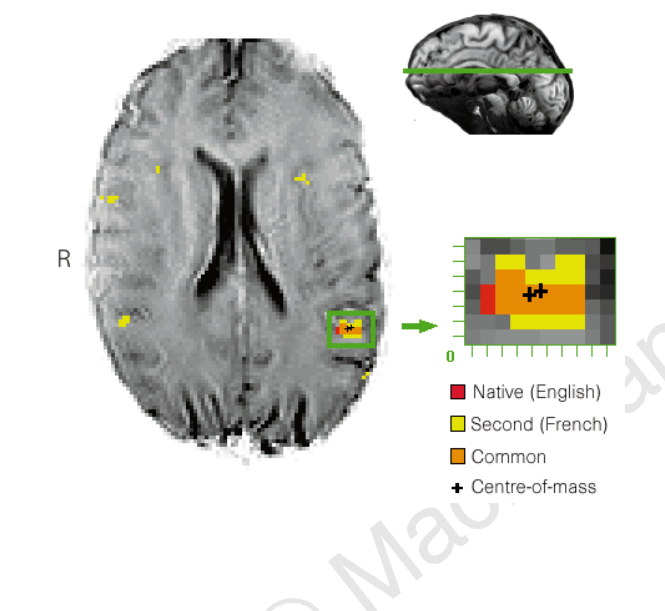


Figure 2 Similar to Fig. 1, an expanded view of the pattern of activity within the superior temporal gyrus (Brodmann's area 22 (refs 2, 3, 18), corresponding to Wernicke's area¹⁻³) indicates centroids of activity for the two languages in this posterior language region. Orange indicates that the voxels that passed all statistical criteria during both the native and acquired language tasks. Centre-of-mass calculations indicate that the centroids are separated on this plane by 1.1 mm, less than the diameter of a single voxel.

language areas was 6.43 (± 1.83) mm and exceeded that of the posterior language areas, which was 1.88 (± 0.62) mm, for these subjects ($t = 5.43$, d.f. = 5, $P \leq 0.004$).

The overall stability of the centre-of-mass centroids with variation in the level of statistical stringency (probability of a false-positive result, P) is illustrated for subject (A) in Fig. 4. In the case of the anterior language area, the centre-to-centre distance between the centroids of activity associated with each language remained within the range of 7.8 to 9.1 mm over statistical stringency levels from $P \leq 0.0002$ to $P \leq 0.02$. In the case of the posterior language area, the centroids remained within the approximate width of one voxel, 1.6 mm, over the same range of stringency levels. Similar results were obtained for all subjects, confirming that threshold criteria do not account for the centre-to-centre distances between centroids of activity associated with each language task.

Figure 5 illustrates the main findings for a typical 'early' bilingual subject (subject (G)) for whom the centre-to-centre distance between the activity centroids (+) of the two activity patterns was 2.3 mm, less than 1.5 voxels (Table 2a). This represents the general pattern for all six early bilingual subjects, where the mean separation

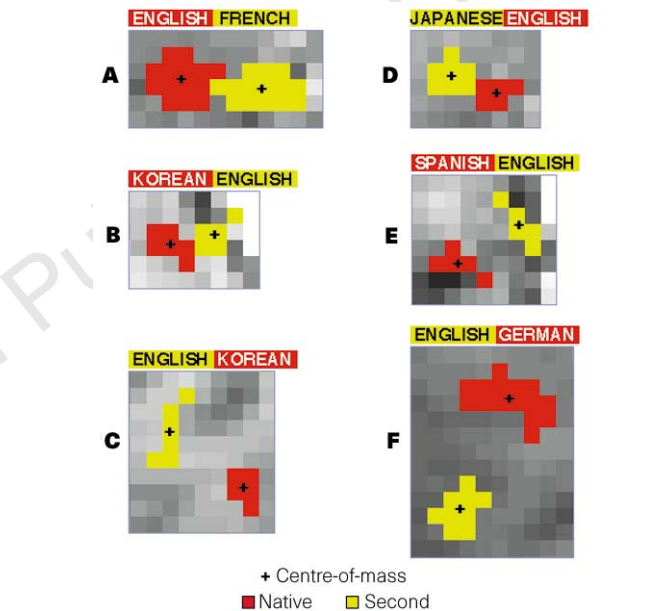


Figure 3 Expanded views of the activity patterns within Brodmann's area 44 (and 46 (refs 2, 3, 18), subject B) for each 'late' bilingual subjects (A-F) indicate the active regions during the native language task (red) and the second acquired language task (yellow). The level of statistical stringency (probability of a false positive result, P) was ≤ 0.0005 for each subject. The centre-of-mass is indicated by a plus sign for each language; area and centre-to-centre (C-C) distances are listed in Table 2a.

was 1.53 (± 0.78) mm. In the case of the posterior language area, Wernicke's area, the mean separation for all early bilingual subjects was 1.58 (± 0.79) mm, similar to the anterior area.

Our findings are summarized by an analysis of variance (Table 3) in which language area (Broca's and Wernicke's) was compared with bilingual type (early and late) with respect to the centre-to-centre distance in millimetres between the two language centroids. Significant main effects for language area ($P \leq 0.000059$) and bilingual type ($P \leq 0.000084$) with an interaction effect ($P \leq 0.000067$) show that activation sites for the two different languages tend to be spatially distinct in Broca's area when the second language was obtained late in life and not when acquired in early childhood; and

Table 3 ANOVA

Source of variation	SS	d.f.	MS	F	P
A (language area)	52.01	1	52.01	25.70	0.000059
B (bilingual type)	48.80	1	48.80	24.11	0.000084
AB	50.79	1	50.79	25.10	0.000067
Within cell	40.47	20	2.02		

Table 2 Areas of activation

(a) Anterior language area											
Subject	Bilingual type	Hemisphere	Gyrus	Brodmann's area	Talairach & Tournoux sectors ^{18*}			Area L1† (mm ²)	Area L2† (mm ²)	C-C distance‡ (mm)	Common voxels L1 and L2† (%)
					X	Y	Z				
A	Late	Left	Inferior frontal	44	d	D	12	34.2	34.2	7.9	0
B	Late	Left	Inferior frontal	44, 46	c	D, C	12	14.6	12.2	4.5	0
C	Late	Left	Inferior frontal	44	c	D	12	12.2	14.6	9.0	0
C'	Late	Left	Inferior frontal	44	c	D	12	15.4	24.2	7.5	0
D	Late	Left	Inferior frontal	44	c	B	1	12.2	17.1	4.7	0
E	Late	Left	Inferior frontal	44	c	D	12	12.2	12.2	7.0	0
F	Late	Left	Inferior frontal	44	c	D	12	33.1	19.8	11.2	0
G	Early	Left	Inferior frontal	44	c	D	12	107.4	75.7	2.3	50
H	Early	Left	Inferior frontal	44	c	D	12	12.2	7.3	0.5	60
I	Early	Left	Inferior frontal	44	c	B	1	41.9	28.6	0.7	46
J	Early	Left	Inferior frontal	44	c	D	12	8.8	4.4	1.7	20
K	Early	Left	Inferior frontal	44	c	D	12	123.4	154.2	2.2	45
L	Early	Left	Inferior frontal	44	c	D	12	44.1	17.6	1.8	22
(b) Posterior language area											
A	Late	Left	Superior temporal	22	d	G	20	39.1	65.9	1.1	48
B	Late	Left	Superior temporal	22	d	G	20	9.8	19.5	1.5	33
C	Late	Left	Superior temporal	22	d	F	8	31.7	34.2	2.2	35
C'	Late	Left	Superior temporal	22	d	G	20	19.8	33.1	3.3	26
D	Late	Left	Superior temporal	22	d	E	1	43.9	19.5	2.2	30
E	Late	Left	Superior temporal	22	d	F	8	22.0	9.8	1.5	18
F	Late	Left	Superior temporal	22	d	G	16	83.7	46.3	1.0	18
G	Early	Left	Superior temporal	22	d	G	16	4.9	31.7	1.7	15
H	Early	Left	Superior temporal	22	d	F	16	56.2	39.1	1.5	56
I	Early	Left	Superior temporal	22	d	E	4	94.7	22.0	3.0	23
J	Early	Left	Superior temporal	22	d	F	16	15.4	125.6	1.4	12
K	Early	Left	Superior temporal	22	d	G	20	121.2	96.9	1.3	62
L	Early	Left	Superior temporal	22	d	F	12	48.5	88.1	0.6	48

C' is a replication of C after 16 months on a different scanner. All observations of area and distance are made at a common level of significance ($P \leq 0.0005$).
^{*} See ref. 18 where the three-dimensional proportional grid system is described. Columns X, Y and Z indicate coordinates of an orthogonal parallelogram, the dimensions of which vary with the principal axes of the brain. Each of these volumes is defined by its three dimensions indicated by a capital letter (Y), a lower-case letter (X), and a number (Z).
[†] L1 and L2 refer to the native and second acquired languages for late bilingual subjects, respectively, and the two languages acquired during the early developmental period for the early bilingual subjects in the respective order as listed on Table 1.
[‡] C-C distance refers to the centre-to-centre spacing between the centre-of-mass centroids of significant voxels associated with the two languages.

that Wernicke's area showed little or no separation of activity regardless of age of acquisition.

The observation that the anatomical separation of the two languages in Broca's area varies with the time at which the second language was acquired, suggests that age of language acquisition may be a significant factor in determining the functional organization of this area in human brain. Human infants, initially capable of discriminating all phonetically 'relevant' differences, may eventually modify the perceptual acoustic space, based on early and repeated exposure to their native languages¹³. It is possible that representations of languages in Broca's area that are developed by exposure early in life are not subsequently modified. This could necessitate the utilization of adjacent cortical areas for the second language learned as an adult.

The difference between the results of this investigation and a positron emission tomography (PET) study in which multiple languages were found to generate overlapping regions of activation within the inferior frontal gyrus¹⁴ may be reconciled in part by the higher effective resolution of this fMRI technique. The intrinsic resolution of the PET H₂¹⁵O cerebral-blood-flow technique was $5 \times 5 \times 6 \text{ mm}^3$ and the results from several subjects were combined and averaged. Individual variability in both the locations of the language areas and the diversity of brain shapes and gyral patterns of the subjects averaged together further reduce the effective resolution of this approach^{9,15}, which could account for the discrepancy. However, on the basis of our findings, the distinction between native and second languages may be less for younger ages of exposure to a second language. The average age of initial exposure to the second language in the PET study was 7.3 years, younger than that of the late bilingual subjects in our study, and could therefore be consistent with our observation for early bilinguals.

To render our findings independent of particular languages or cultural background, our study made use of simple expressive tasks with similar semantic content across multiple languages and

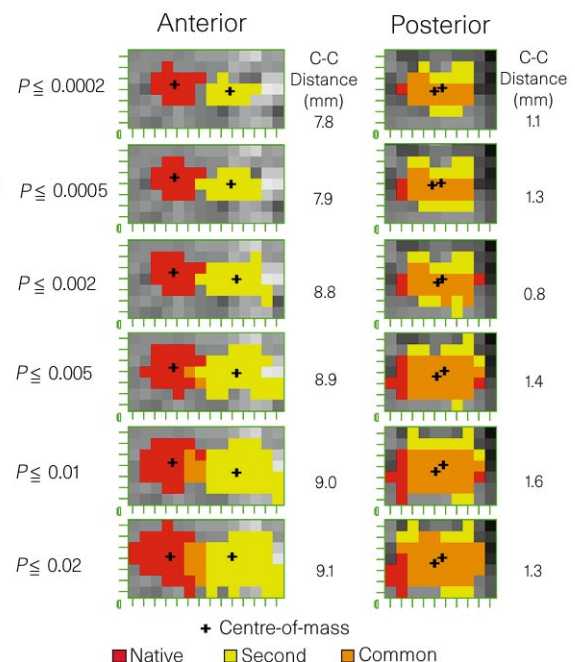


Figure 4 All voxels that pass the statistical criteria during the native (red) and acquired (yellow) language tasks (subject (A)) are shown for the anterior region of interest, left, and the posterior region of interest, right, over approximately 3 orders of statistical stringency ($P \leq 0.0002$ to $P \leq 0.02$). The centre-to-centre (C-C) distance between the centroids of activity in the anterior area (based on the centre-of-mass, +), ranges from 7.8 mm at $P \leq 0.0002$ to 9.1 mm at $P \leq 0.02$, suggesting that the centroid is nearly independent of cluster size (radial expansion). Similarly, for the posterior language area, the C-C distance ranges from 1.1 to 1.3 mm (less than the width of a single voxel) over the same level of statistical stringency.

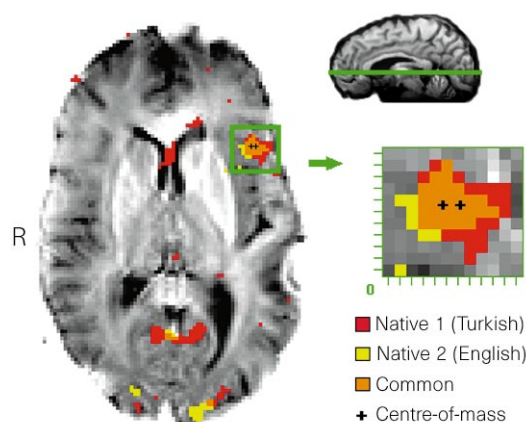


Figure 5 A representative axial slice from an 'early' bilingual subject (G) who learned English and Turkish simultaneously during early childhood shows all voxels that pass the multistage statistical criteria at $P \leq 0.0005$. Red indicates the Turkish language task and yellow indicates the English language task. An expanded view of the region of interest (Broca's area¹⁻³) indicates multiple common voxels between the two language areas. The geometric centres-of-mass indicate that the centroids are within 1.5 voxels. R indicates the right side of the brain.

included subjects with various combinations of languages (Table 1). Our findings are consistent with distinct roles for the anterior and posterior language areas in the processing of human language, and raise further questions regarding the role of Broca's area in processing the phonetic structures of different languages. □

Methods

Imaging. A 1.5-tesla magnetic resonance scanner (General Electric) retrofitted (advanced NMR Instascan) for echo-planar imaging and subsequently upgraded to the GE echo-planar system was used to obtain T2*-weighted images with a gradient echo pulse sequence (echo time, 60 ms; repetition time, 3,000 ms; flip angle, 30°) which is sensitive to magnetic resonance signal changes caused by alteration in the proportion of deoxyhaemoglobin in the local vasculature accompanying neuronal activation¹⁶. Either a volume-optimized 5 × 5-mesh dome resonator¹⁷ or a General Electric head coil was employed. The in-plane resolution was 1.6 mm by 1.6 mm. Slice thickness was 4.5–4.7 mm and 16 contiguous slices of brain were obtained parallel to a reference line through the superior edge of the anterior commissure and the inferior edge of the posterior commissure^{18,19}. These slices covered the inferior frontal gyrus (the anterior language region, 'Broca's' area including Brodmann's areas 44 and 46) and the posterior superior temporal gyrus (the posterior language region, 'Wernicke's' area including Brodmann's area 22)^{1-3,18,19}. Thirty images were taken, one every 3 s; thus, an entire run lasted 90 s. The first 10 images (30 s) were acquired during a baseline period, followed by a stimulation or task period of 10 images (30 s), and a final (30 s) baseline period also consisting of 10 images. A fixation cross-hair was provided during the baseline epochs to help the subject to maintain a stable head position.

Analysis. Two identical runs were performed in each language. Before statistical analysis, all brain images were computationally aligned to allow direct spatial comparisons between different language tasks for individual subjects²⁰, and a two-dimensional gaussian filter (approximately 3 volume elements, voxels, at half-height) was applied to the data. Significant signal changes were identified by a multistage statistical analysis which compared average baseline and stimulation signal intensities and required significant signal changes on two runs (coincidence)^{21,22}. The rate of false-positive voxels, p , was empirically determined from images of a copper sulphate solution-filled spherical phantom (General Electric standard) and found to be less than 0.0005. The centroid of a cluster of language-activated voxels was determined as the two-dimensional centre of mass, and the centre-to-centre distance between centroids was taken as the separation (mm) between language-specific activity.

Task. The sentence-generation task was performed silently (internal speech) to

minimize head movement and was similar to tasks previously employed in neuroimaging language studies²³. The subject was instructed to "describe" events that occurred during a specified period of the previous day (morning, afternoon, night); this task was practised before the imaging sessions. Immediately before each run, the subject was instructed which language he/she was to imagine speaking, and graphical cues signalling morning, afternoon, and night were displayed in various orders for 10 s during the 30-s task period. These graphics provided common non-linguistic cues for the task and the unpredictable order of presentation presumably reduced the tendency to rehearse mentally before the cue. The languages were alternated during the imaging session to prevent habituation and a potentially time-dependent bias.

Subjects. Twelve healthy multilingual volunteers, 9 males and 3 females, were recruited according to institutional informed consent procedures. Subjects were either right-handed or ambidextrous, as assessed by the Edinburgh handedness inventory²⁴ (Table 1). The mean age of subjects was 29.3 (± 4.2) years. Six subjects ('early' bilinguals) were exposed to two languages during infancy, and six subjects ('late' bilinguals) were exposed to a second language in early adulthood. The mean age of initial exposure to the second language was 11.2 (± 1.5) years and the mean age that conversational fluency was achieved was 19.2 (± 4.1) years. Each of the 'late' bilingual subjects had lived in the country of the second language, which assured a high standard for fluency. Each of the early bilinguals was raised in a home where either the parents spoke one language and siblings and friends spoke another, or the parents spoke two languages. Ten languages were represented as indicated on Table 1, and all subjects reported approximately equal fluency and frequent usage in each language at the time of testing.

Received 20 February; accepted 30 April 1997.

1. Geschwind, N. The organization of language and the brain. *Science* **170**, 940–944 (1970).
2. Kretschmann, H.-J. & Weinrich, W. *Cranial Neuroimaging and Clinical Neuroanatomy* 2nd edn (Thieme Medical, New York, 1992).
3. Damasio, H. & Damasio, A. *Lesion Analysis in Neuropsychology* (Oxford Univ. Press, New York, 1989).
4. Gomez-Tortosa, E. *et al.* Selective deficit of one language in a bilingual patient following surgery in the left perisylvian area. *Brain Lang.* **48**, 320–325 (1995).
5. Schwartz, M. S. Ictal language shift in a polyglot. *J. Neurol. Neurosurg. Psychiatr.* **57**, 121 (1994).
6. Ojemann, G. A. Brain organization for language from the perspective of electrical stimulation mapping. *Behav. Brain Sci.* **6**, 189–230 (1983).
7. Black, P. M. & Ronner, S. F. Cortical mapping for defining the limits of tumor resection. *Neurosurgery* **20**, 914–919 (1987).
8. Petsche, H., Etlinger, S. C. & Filz, O. Brain electrical mechanisms of bilingual speech management: an initial investigation. *Electroencephalogr. Clin. Neurophysiol.* **86**, 385–394 (1993).
9. Démonet, J. F., Wise, R. & Frackowiak, R. S. Language functions explored in normal subjects by positron emission tomography: A critical review. *Human Brain Mapping* **1**, 39–47 (1993).
10. Zatorre, R. On the representation of multiple languages in the brain: Old problems and new directions. *Brain and Language* **36**, 127–147 (1989).
11. Ojemann, G. A. Cortical organization of language. *J. Neurosci.* **11**, 2281–2287 (1991).
12. Zatorre, R. J., Meyer, E., Gjedde, A. & Evans, A. C. PET studies of phonetic processing of speech: Review, replication, and reanalysis. *Cereb. Cort.* **6**, 21–30 (1996).
13. Kuhl, P. K. Learning and representation in speech and language. *Curr. Opin. Neurobiol.* **4**, 812–822 (1994).
14. Klein, D. *et al.* The neural substrates underlying word generation: A bilingual functional-imaging study. *Proc. Natl Acad. Sci. USA* **92**, 2899–2903 (1995).
15. Steinmetz, H. & Seitz, R. J. Functional anatomy of language processing: Neuroimaging and the problem of individual variability. *Neuropsychologia* **29**, 1149–1161 (1991).
16. Ogawa, S., Lee, T.-M., Nayak, A. S. & Glynn, P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Med.* **14**, 68–78 (1990).
17. Meyer, K. L. *et al.* Sensitivity-enhanced echo-planar MRI at 1.5 T using a 5 × 5 mesh dome resonator. *Magn. Reson. Med.* **36**, 606–612 (1996).
18. Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical, New York, 1988).
19. Damasio, H. *Human Brain Anatomy in Computerized Images* (Oxford Univ. Press, New York, 1995).
20. Woods, R. P., Mazziotta, J. C. & Cherry, S. R. MRI-PET registration with automated algorithm. *J. Comput. Assist. Tomogr.* **17**, 536–546 (1993).
21. Hirsch, J. *et al.* A multi-stage statistical technique to identify cortical activation using functional MRI. *Proc. Soc. Mag. Res.* **2**, 637 (1994).
22. Hirsch, J. *et al.* Illusory contours activate specific regions in human visual cortex: Evidence from functional magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **92**, 6469–6473 (1995).
23. Hinkle, R. M. *et al.* Functional magnetic resonance imaging of Broca's area during internal speech. *NeuroReport* **4**, 675–678 (1993).
24. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).

Acknowledgements. We thank K. Zakian and D. Ballon for use of the 5 × 5 mesh dome resonator, J. Victor, G. Krol, J. Posner, R. Cappiello, M. Ruge, D. Correa, S. Harris, J. Salvagno, P. Kuhl, F. Nottebohm, G. E. Vates and R. DeLePaz for technical assistance and helpful comments, and Y. Popowich, N. Rubin, T. Ozaki, D. R. Moreno, B. Aghazadeh, D. Barbut-Heinemann, D. Orbach, R. Valencia, J. Carton, E. Götte, R. Härtl, O. Torres and M. Li for volunteering as subjects. Supported by the William T. Morris Foundation fellowship, the Tri-Institutional MD/Ph.D Program (KHSC); the Charles A. Dana Foundation, Johnson & Johnson Focused Giving Foundation, Cancer Center Support Grant NCI (J.H.); the C. V. Starr Foundation and the Lookout Fund (N.R.R.).

Correspondence and requests for materials should be addressed to J. Hirsch (e-mail: hirsch@vision.mskcc.org).

Representation of motion boundaries in retinotopic human visual cortical areas

John B. Reppas^{*†}, Sourabh Niyogi[‡], Anders M. Dale[†], Martin I. Sereno[§] & Roger B. H. Tootell[†]

^{*} Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

[†] Massachusetts General Hospital Nuclear Magnetic Resonance Center, 149 13th Street, Charlestown, Massachusetts 02129, USA

[‡] Perceptual Science Group, Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[§] Department of Cognitive Sciences, University of California at San Diego, La Jolla, California 92093-0515, USA

Edges are important in the interpretation of the retinal image. Although luminance edges have been studied extensively, much less is known about how or where the primate visual system detects boundaries defined by differences in surface properties such as texture, motion or binocular disparity. Here we use functional magnetic resonance imaging (fMRI) to localize human visual cortical activity related to the processing of one such higher-order edge type: motion boundaries. We describe a robust fMRI signal that is selective for motion segmentation. This boundary-specific signal is present, and retinotopically organized, within early visual areas, beginning in the primary visual cortex (area V1). Surprisingly, it is largely absent from the motion-selective area MT/V5 and far extrastriate visual areas. Changes in the surface velocity defining the motion boundaries affect the strength of the fMRI signal. In parallel psychophysical experiments, the perceptual salience of the boundaries shows a similar dependence on surface velocity. These results demonstrate that information for segmenting scenes by relative motion is represented as early as V1.

Clinical and psychophysical evidence suggests that humans possess specialized mechanisms for detecting motion boundaries^{1,2}. However, it is unclear where boundary detection occurs in the visual pathway. The observation that motion segmentation is rapid and pre-attentive suggests an early visual mechanism, and there is some neurophysiological evidence for this⁴. At the same time, boundary mechanisms also integrate information over a large region of visual space and a range of motion coherences^{5,6}. These latter properties are consistent with an extrastriate detection mechanism—a view supported by theoretical considerations⁷ and by preliminary human brain-mapping results^{8,9}. Here we test whether an initial motion-sensing stage in early visual areas (for example, V1, V2, MT) precedes motion boundary detection by later areas in the human dorsal or ventral visual processing streams.

The stimulus was a full-field random dot texture pattern that alternated between periods of segmented and uniform motions. In both the uniform and segmented conditions, the direction of local motion reversed every 1.2 Hz. Thus, the time-averaged local image statistics (contrast, spatial pattern, direction of motion) were identical for the segmented and uniform portions of the scan. The two conditions were distinguished only by the presence or absence of motion boundaries. We used boundaries created in three ways: (1) by shearing motions, in which local motions were parallel to the motion edge; (2) by compressive motions, in which local motions were perpendicular to the edge; and (3) by a checkerboard boundary pattern containing both types of edges. The default texture velocity was 5 degrees per second. This velocity produced the most visible boundaries in our displays, and is typical of velocities used in perceptual studies of motion segmentation^{6,9}.

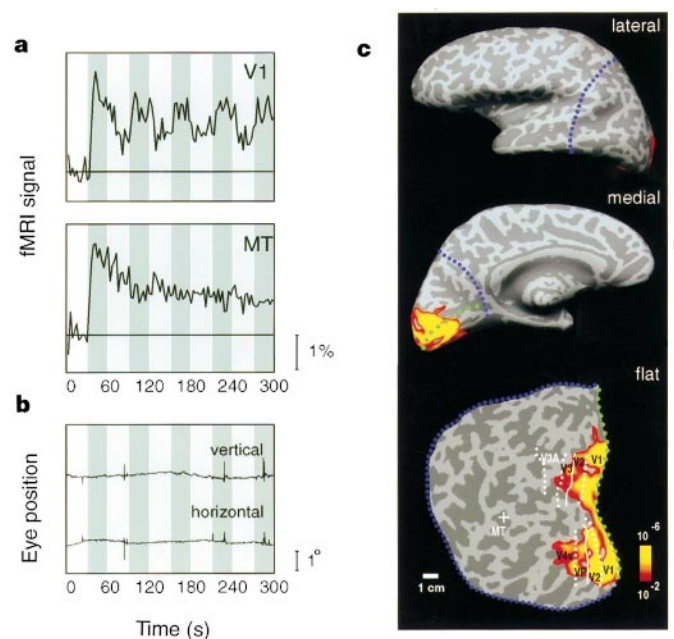


Figure 1 The motion segmentation fMRI signal and its anatomical distribution. The subject viewed a uniform motion (light green shading) alternating every 30 s with a checkerboard pattern formed by relative motion (darker shading). One 'check' subtended ~ 0.9 by 0.9 degrees. For the first 30 s, the display was blank except for the fixation point. **a**, The fMRI signal changes in V1 and MT/V5 are expressed as a percentage change from the signal during this baseline period. **b**, Horizontal and vertical positions of the right eye; shading as in **a**. Transients correspond to blinks. **c**, Areas selectively activated during the motion boundary portions of the scan. Dotted blue and green lines on the medial and lateral views of the inflated hemisphere indicate the 'cuts' needed to flatten the occipital lobe. Gyri and sulci are indicated in light and dark grey, respectively. Solid and dotted white lines indicate the representation of the horizontal and vertical meridia respectively (see Methods). Scale bar, 1 cm.

Moreover, it is well matched to the velocity selectivity reported for opponent motion neurons in the non-human primate¹⁰. Subjects were instructed to fixate a central cross-hair and view the stimulus passively. In some cases, eye movements were monitored to document the stability of fixation.

Compared to a baseline of mean luminance, the response to the experimental stimulus had two components in area V1 and the other retinotopic areas: (1) a constant elevation of the fMRI signal beginning at the onset of the stimulus and lasting the entire experiment; and (2) a stimulus-locked response superimposed on the constant response (Fig. 1a). The constant component represents the response to the visual stimulus elements that are present throughout the experiment, namely the visual texture and its reversing motion. The stimulus-locked component originates from mechanisms that are selectively activated during segmented but not uniform motion. Eye movements were not significantly different during segmented versus uniform motion periods (Fig. 1b), suggesting that the modulated fMRI signal seen in the retinotopic visual areas is due to interactions of the stimulus with purely sensory mechanisms. In area MT/V5 and far extrastriate visual areas, the fMRI response typically consisted only of the sustained component. We never observed an elevated response for uniform over segmented motions in any part of the visual cortex. The fMRI signals did not depend critically on the type of motion boundary used, whether shearing, compressive or both.

The motion segmentation signal was localized on the flattened visual cortices of 6 subjects (12 hemispheres). In each subject we also mapped the boundaries of the overtly retinotopic visual areas

(V1, V2, V3/VP, V3A and V4v) using phase-encoded retinotopic stimulation^{11–13}. Area MT/V5 was located using a low-contrast radial motion stimulus¹⁴. Figure 1c shows the response to a motion-boundary defined checkerboard pattern in a representative subject relative to the retinotopic visual areas and MT/V5. For the high density of motion boundaries in this stimulus, the selective response to motion boundaries is strongest in V1 and V2, present in V3 and VP and largely absent in V3A and V4v. Strong fMRI signals selective for motion boundaries were not observed in MT/V5 of this subject, nor in those regions of the occipital, temporal and parietal lobes which are thought to contain the human homologues of macaque dorsal and ventral visual streams¹⁵. Similar results were obtained in the five other subjects.

When the number of motion boundaries in the stimulus was reduced, we observed that the segmentation signal within the early visual areas became patchy, suggestive of discrete responses to individual boundaries within the visual field. To test this hypothesis, we presented a single motion boundary in a region of the visual field that could be identified easily using an independent retinotopic criterion. The stimulus was a rotating texture pattern which alternated between a uniform motion and a shearing motion that created a single, stationary, circular boundary located ~7 degrees eccentric to the fixation point (Fig. 2a). The representation of this circle in visual space was independently confirmed from isoeccentricity phase maps in the same subjects¹².

The circular motion boundary produced a modulated response in two distinct regions (Fig. 2a). These occur in the medial bank of the occipital lobe and are oriented perpendicular to the boundaries between retinotopic visual areas. The stripes of activation are limited to V1, V2 and V3/VP and occur at the representation of the 7-deg isoeccentricity line. The spatial extent of the boundary activation is comparable to that produced by a thin ring of counter-phasing checkerboard in the same position. This suggests that the V1/V2/V3 response to motion boundaries engages a detection mechanism of similar dimensions to that of luminance. Similar results were obtained with stationary wedge boundaries formed by compressive dot motions. If the single boundary moved through the visual field during the segmented epochs, the localized modulation of the fMRI signal was no longer observed. This result demonstrates that motion boundaries not only activate early visual areas, but that they do so in a retinotopically specific fashion.

We examined the strength of the fMRI segmentation signal as the

number of boundaries in the display was systematically varied. The stimulus for this experiment alternated between a rotating uniform motion and a pinwheel pattern formed by compressive motions (5% dot density). The pinwheel pattern was chosen because texture velocity and boundary size both vary with eccentricity. Because motion-defined boundaries are detected by a localized retinotopic mechanism (Fig. 2), the motion edges were rotated with the carrier motion to engage detection mechanisms at all spatial locations. Five subjects were scanned using the standard velocity of 5 degrees per second, where the boundaries are most visible (Fig. 4). The tuning curves were repeated in two subjects at a higher velocity (30 degrees per second), better matched to the median velocity selectivity of MT cells¹⁶. We wanted to eliminate the possibility that the lack of a modulated signal in MT/V5 simply reflected the relatively low velocity of the standard carrier motion. At these high carrier velocities, subjects report not seeing any motion boundaries (Fig. 4c).

Figure 3 summarizes how the segmentation fMRI signal varied with the number of boundaries. For the retinotopic areas, the curves have an inverted 'U' shape: the modulated response is low for sparse and dense boundary configurations, and maximal for an intermediate density. The peaks of these tuning curves are different for each retinotopic area. V1, which represents one extreme of the response, is optimally stimulated by motions that are segmented at a fine scale (16–32 wedges in the display), whereas V3A, at the other extreme, is best stimulated by coarser motion segmentation (4–8 wedges in the display). The peaks of areas V2 and V3/VP are intermediate between V1 and V3A (results not shown). These results suggest that spatial scale is an important determinant of the fMRI response to segmented motions, and that one function of the retinotopic areas may be to segment moving images at different scales¹⁷. Data from MT/V5 show that it is much less selectively activated than the retinotopic areas over the entire range of boundary densities, for both high and low velocity displays. When stationary boundaries were used, the results were similar (data not shown). For dense texture displays, the tuning curves extended to higher boundary densities. Unlike single units in macaque MT, no suppression of the MT/V5 fMRI signal was observed for transparent motions¹⁸, but the stimulus and physiological methods in the present experiments differed from those previous experiments.

The response curve for each retinotopic area falls off at boundary edges when subjects reported no longer seeing distinct kinetic edges, but rather a display that approximated transparent motion¹⁹.

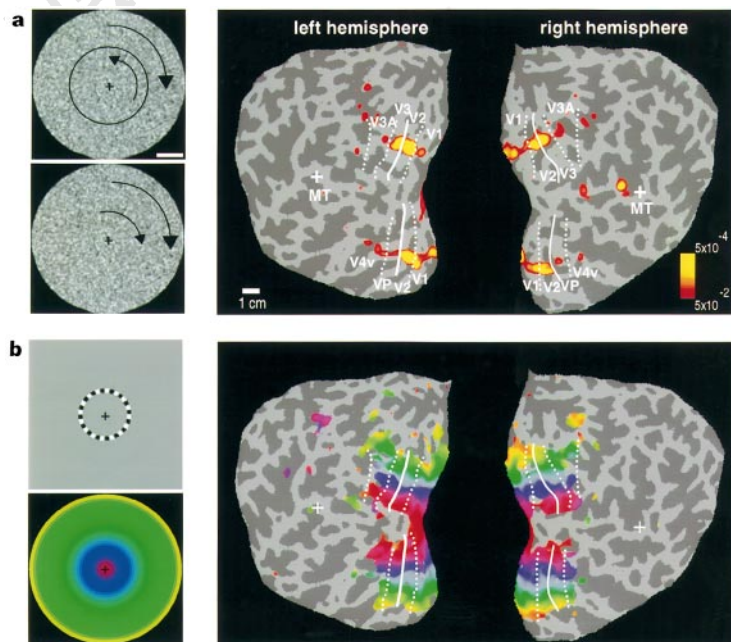


Figure 2 A circular motion boundary produces modulated fMRI activity at the cortical representation of that circle. **a**, Left, the subject viewed a single circular motion boundary about 7 deg eccentric to the fixation point which alternated with uniform motion; a single frame of this texture is shown (scale bar, 5 deg). Right, the yellow to red pseudocolour indicates the distribution of stimulus-locked modulation in the fMRI signal in the left and right flattened hemispheres. The borders of the visual areas are as in Fig. 1. **b**, Isoeccentricity maps¹² for a thin checkerboard ring (top left panel) which expanded repeatedly from the fixation point to the visual periphery. Pseudocolour (right) indicates eccentricity in the visual field as determined by phase-encoded retinotopic mapping. The pattern of localized activity in **a** corresponds to a region of constant eccentricity in panel **b**. Scale bar, 1 cm.

In fact, we observed no selective fMRI modulation in experiments of transparent versus uniform motions (Fig. 3). This finding provides an important control for the hypothesis that the fMRI signal we observed is a bona fide boundary signal. The presence of two directions of motion in the same part of the visual field, which is the case for both motion boundaries and motion transparency, is not sufficient to generate a selective fMRI response. A perceptual boundary must be present.

To explore further the origin of the modulated fMRI signal, we examined how motion boundaries formed by different surface velocities affect the motion segmentation signal. Boundaries formed by two very slow or two very fast motions are harder to see than those formed by two intermediate motion velocities, at least for brief stimulus presentation^{20,21}. If the modulated response we observe originates from a boundary detection mechanism, it

should be: (1) related to the subjective experience of motion boundaries over a range of boundary visibility; and (2) dissociable from a pure visual motion signal. The fMRI segmentation signal in V1 was measured for different surface velocities with a boundary density (~ 0.5 cycles per degree (c.p.d.)) optimized for V1 fMRI responses in four subjects. In separate tests, we measured the visibility of these boundaries as a function of surface velocity using a two-alternative forced-choice task in which subjects were instructed to discriminate segmented from uniform motions.

The modulated component of the V1 fMRI signal is maximal at an intermediate velocity of ~ 5 degrees per second, and decreases for higher and lower velocities (Fig. 4a). Because the sustained component of the fMRI signal does not vary with stimulus velocity, this variation in the strength of the segmentation signal cannot be explained by the failure of V1 to register fast and slow motions

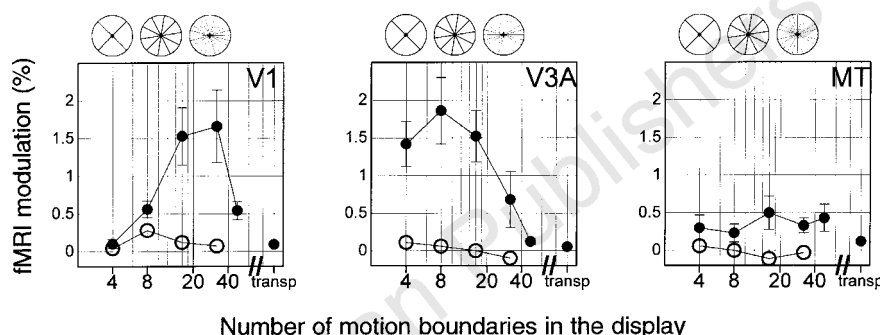


Figure 3 Tuning of the fMRI segmentation signal in three visual areas as a function of boundary density. The stimulus was a rotating texture (dot density, 5%) that alternated every 30 s between uniform rotational motion and pinwheel boundaries formed by compressive motions. Wedge angle varied from 90 deg to 7.5 deg of arc. We measured the strength of the fMRI segmentation signal as the number of wedges in the stimulus was varied. A given density was shown 1–3 times in a randomly interleaved fashion, and the results were averaged for each density tested. Filled points represent the mean $\pm$ s.d. percentage modulation of

the fMRI signal for 5 subjects at 5 deg s^{-1} . Open points are the mean percentage modulation for 2 subjects at 30 deg s^{-1} . At low velocity, retinotopic areas showed robust responses to the boundary stimulus and were selective for a range of boundary densities. However, when transparent motion was alternated with uniform motion, they showed no selective activation. MT/V5's response to boundaries was less prominent and not tuned for boundary density at either velocity.

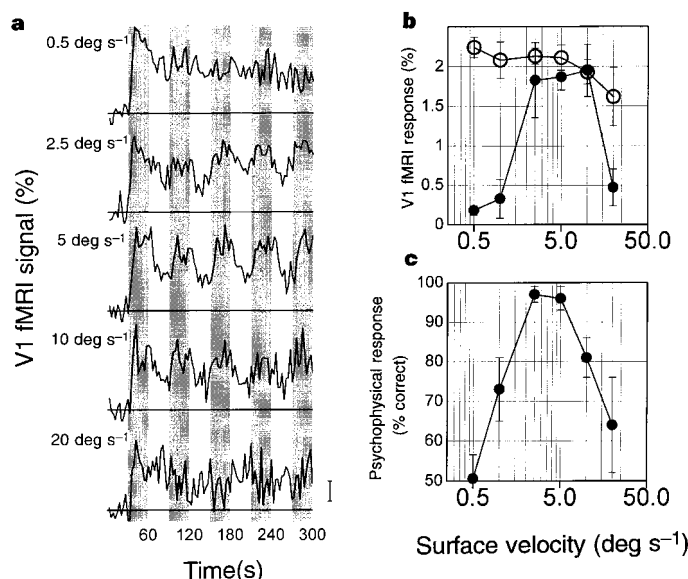


Figure 4 The modulated component of the fMRI segmentation signal varies with the perceptual salience of the motion boundary display. The stimulus, optimized for area V1, contained a fixed number of concentric circular boundaries (~ 0.5 c.p.d.) formed by shearing motion of the dense texture. These alternated with uniform rotational motion every 30 s. In successive scans, the velocity of the carrier motion was varied. Each of the six velocities was presented 1–3 times per scanning session in a randomly interleaved fashion. **a**, Raw time-courses for 5

velocities, sampled from V1 of one subject. Shading indicates periods of segmented motion. Scale bar, 1% signal change. **b**, Averaged strength of both the modulated (black circles) and steady-state (white circles) signal change relative to the mean luminance baseline (mean $\pm$ s.d.; $n = 4$). **c**, Mean performance on a psychophysical task that required subjects to distinguish segmented from uniform motion (mean $\pm$ s.d.; $n = 3$).

per se (Fig. 4b). Rather, the average modulated fMRI response is related to the visibility of the boundaries, expressed as average performance in a standard boundary detection task²¹. As subjects see the boundaries less distinctly, the fMRI signal in V1 modulates less and less in response to the stimulus (Fig. 4c). The duration of the stimuli used in the fMRI experiment is much longer than that used for psychophysical detection experiments, and at these longer durations the boundary is easily perceived at all velocities tested. Thus the fMRI signal does not correlate directly with perception. Nevertheless, the similarity in the effects of velocity in the two experiments (compare Fig. 4b,c) suggests that the fMRI signal reflects processes within V1 that are likely also to be relevant to perception of motion boundaries. The dissociation between the fMRI response to the boundaries themselves versus the motions that define the boundaries suggests that the response properties of motion boundary mechanisms in V1 differ from the better-known striate cortical mechanisms that register stimulus pattern and motion.

Our results are inconsistent with the views that: (1) neurons in early visual areas (V1, V2, V3) register only the local direction of image motion; and (2) selectivity for higher-order optic flow patterns, like motion boundaries, is developed only late in the motion processing pathway (for example, MT and MST)⁶. The fMRI responses we describe may originate from boundary-sensitive mechanisms within the retinotopic areas, possibly involving long-range horizontal connections⁴. Alternatively, feedback from extrastriate motion areas may contribute to the segmentation responses seen in retinotopic areas²².

Our findings do not preclude the contribution of non-retinotopic cortical areas, and area MT/V5 in particular, to the detection of motion boundaries. The lack of a strongly selective motion segmentation signal from MT/V5 over a range of boundary densities and surface velocities might simply result from the presence of approximately equal numbers of wide-field and opponent-motion neurons²³. Alternatively, human MT/V5 may simply not be important for motion segmentation, consistent with a previous PET study⁹, neuropsychological findings⁵, and experimental results in the macaque^{24,25}. Certainly one important role of monkey MT appears to be the integration of local motion signals across space and time to form global representations of motion²⁶—a process that may 'blur' motion boundary information. □

Methods

Experimental design. All imaging experiments compared fMRI signals during a motion boundary condition to those produced during a uniform motion condition. These conditions alternated either every 16 s (Figs 1c, 2) or every 30 s (Figs 1a, 3, 4). The order of segmented and uniform conditions varied. Boundaries were not visible in the absence of display motion, and single stimulus frames from the boundary condition contained no significant Fourier energy at the orientations and bandwidths of the motion boundary pattern. We used two types of random dot stimuli: a dense version, in which each pixel was assigned to one of 255 grey values, and a sparse pattern of white dots on a black background (dot size $\sim 0.15 \times 0.15$ deg; overall dot density was 5%). The dot velocity was ~ 5 deg s⁻¹ unless stated otherwise; this is the display velocity at which boundaries are most visible (Fig. 4). The stimuli were generated on a Silicon Graphics Onyx computer using GL graphics functions at a frame rate of 60 Hz. The output of the graphics window was sent to a Sony 2000 LCD projector, which projected images into the bore of the magnet, where they subtended ~ 26 by 26 degrees. Software control of screen intensity was calibrated using a Spectraspot photometer (Photoresearch). Eye position was recorded using an infrared monitoring device (Ober2, Sweden) and sampled at 10 Hz. Linear drift of the eye position traces over the recording period was subtracted. In all, 16 subjects were scanned with motion boundary stimuli. Of these, six were studied in detail using cortical flat mapping techniques, six more contributed to tuning curves (Figs 3, 4), and data from the remainder were used to confirm the basic findings. Most subjects used rigidly mounted, individually moulded, deep-impression bite bars. Head motion did not affect scans that

contributed to the flat maps and tuning curves shown, presumably reflecting the constraints of the bite bar and the previous fMRI experience of these subjects.

fMRI scans. Subjects were scanned on a 1.5 T GE magnet, retrofitted for echo-planar imaging (Advanced NMR), using a bilateral quadrature surface coil overlying the occipital, inferior parietal and posterior temporal cortex. In functional imaging experiments, we used an asymmetric spin-echo sequence (TR = 2,000–4,000 ms, TE = 70 ms, with a 180 deg refocusing pulse offset of -25 ms), to visualize the BOLD (blood oxygen level-dependent) contrast changes that accompany brain activity^{14,27–29}. Images were acquired in 14–16 contiguous 4-mm slices oriented approximately perpendicular to the calcarine fissure; in-plane resolution was 3 mm $\times$ 3 mm. Statistical maps of the modulated fMRI signal were constructed by performing a fast-Fourier transform on each voxel's time-course and performing an F-test on the ratio of stimulus-locked power to all other non-harmonic frequencies, for frequencies above 3 cycles per scan. The resulting *P* values (null hypothesis: no stimulus-locked modulation) were smoothed using 10 iterations of a box-car filter and thresholded. Three-dimensional reconstructions of the subjects' brains were 'inflated' by an iterative algorithm that relaxed and minimized angular and areal distortion³⁰. Flattened views of the occipital lobe were obtained by removing the occipital region from the inflated hemispheres, cutting it open along the fundus of the calcarine fissure, and allowing it to unfold¹². The boundaries between the overtly retinotopic areas were determined by transitions in the visual field sign map, which is derived from a combination of eccentricity and polar angle phase maps^{11–13}. Here, the representations of the vertical meridian are dotted, and the horizontal meridian are solid. MT/V5 was localized by its selective response to a moving versus stationary concentric ring pattern of low (2–3%) contrast¹⁴. The boundaries between retinotopic areas and the centroid of the MT activation were overlaid onto maps of the motion segmentation fMRI signal. For those experiments in which we acquired time-course data, all voxels that were 2–3 mm (one voxel) within the boundaries of each of the labelled areas and which were not obviously confounded by macrovascular artefact (high amplitude/frequency in the time course, and/or location coextensive with large vessels in the angiogram) were selected and averaged together¹⁴. Stimulus-evoked modulation is expressed as the percentage excursion from the signal in the unstimulated condition.

Psychophysics. We used a temporal two-alternative forced choice task requiring subjects to distinguish between the concentric motion boundaries used in the fMRI experiments (Fig. 4) and a uniform motion at different carrier speeds. Each trial began with a blank screen with a central fixation point, and contained two 100-ms stimulus presentations (one segmented and one uniform condition) separated by ~ 1 s (ref. 21). The order of the segmented and uniform motion conditions, the directions of the uniform motion and the overall stimulus speed were all randomly varied from trial to trial. Subjects were instructed to indicate whether the boundary stimulus came first or second in the trial, and to guess when unsure. Performance was assessed by the percentage of correct responses (30 trials per velocity).

Received 19 December 1996; accepted 30 April 1997.

1. Braddick, O. A short-range process in apparent motion. *Vision Res.* **14**, 519–527 (1974).
2. Regan, D., Giaschi, D., Sharpe, J. A. & Hong, X. H. Visual processing of motion-defined form: selective failure in patients with parietotemporal lesions. *J. Neurosci.* **12**, 2198–2210 (1992).
3. Dick, M., Ullman, S. & Sagi, D. Parallel and serial processes in motion detection. *Science* **237**, 400–402 (1987).
4. Lamme, V. A., van Dijk, B. W. & Spekreijse, H. Contour from motion processing occurs in primary visual cortex. *Nature* **363**, 541–543 (1993).
5. Vaina, L. M., Grzywacz, N. M. & Kikinis, R. Segregation of computations underlying perception of motion discontinuity and coherence. *Neuroreport* **5**, 2289–2294 (1994).
6. Murakami, I. & Shimojo, S. Modulation of motion aftereffect by surround motion and its dependence on stimulus size and eccentricity. *Vision Res.* **35**, 1835–1844 (1995).
7. Nakayama, K. & Loomis, J. M. Optimal velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. *Perception* **3**, 63–80 (1974).
8. Patzwahl, D. R., Zanker, J. M. & Altenmüller, E. O. Cortical potentials reflecting motion processing in humans. *Vis. Neurosci.* **11**, 1135–1147 (1994).
9. Orban, G. A. et al. A motion area in human visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 993–997 (1995).
10. Lagae, L., Gulyas, S., Raiguel, S. & Orban, G. A. Laminar analysis of motion information processing in macaque V5. *Brain Res.* **496**, 361–367 (1989).
11. Engel, S. A. et al. fMRI of human visual cortex. *Nature* **369**, 525 (1994).
12. Sereno, M. I. et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).
13. DeYoe, E. A. et al. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl Acad. Sci. USA* **93**, 2382–2386 (1996).
14. Tootell, R. B. H. et al. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).

15. Reppas, J. B., Dale, A. M., Sereno, M. I. & Tootell, R. B. H. The vision: a perception subjective. *La Recherche* **289**, 52–56 (1996).
16. Maunsell, J. H. R. & van Essen, D. C. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* **49**, 1127–1147 (1983).
17. Braddick, O. Segmentation versus integration in visual motion processing. *Trends Neurosci.* **16**, 263–268 (1993).
18. Snowden, R. J., Treue, S., Erickson, R. E. & Andersen, R. A. The response of area MT and V1 neurons to transparent motion. *J. Neurosci.* **11**, 2768–2785 (1991).
19. van Doorn, A. J. & Koenderink, J. J. Spatial properties of the visual detectability of moving spatial white noise. *Exp. Brain Res.* **45**, 189–195 (1982).
20. Banton, T. & Levi, D. M. The perceived strength of motion-defined edges. *Perception* **22**, 1195–1204 (1993).
21. Sachtler, W. L. & Zaidi, Q. Visual processing of motion boundaries. *Vision Res.* **35**, 807–826 (1995).
22. Maunsell, J. H. R. & van Essen, D. C. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.* **3**, 2436–2586 (1983).
23. Born, R. T. & Tootell, R. B. H. Segregation of global and local motion processing in primate middle temporal visual area. *Nature* **357**, 497–499 (1992).
24. Schiller, P. H. The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Vis. Neurosci.* **10**, 717–746 (1993).
25. Marcar, V. L., Xiao, D. K., Raiguel, S. E., Maes, H. & Orban, G. A. Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. *J. Neurophysiol.* **74**, 1258–1270 (1995).
26. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* **10**, 1157–1169 (1993).
27. Kwong, K. K. *et al.* Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl Acad. Sci. USA* **89**, 5675–5679 (1992).
28. Ogawa, S. *et al.* Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **89**, 5951–5955 (1992).
29. DeYoe, E. A., Bandettini, P., Neitz, J., Miller, D. & Winans, P. Functional magnetic resonance imaging (fMRI) of the human brain. *J. Neurosci. Meth.* **54**, 171–187 (1994).
30. Dale, A. M. & Sereno, M. I. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J. Cogn. Neurosci.* **5**, 162–176 (1993).

Acknowledgements. We thank E. Adelson, S. Brown, M. Livingstone and W. van Duffel for comments on earlier versions of this manuscript. S. Macknik kindly loaned us the eye-tracking device used in some experiments. We are grateful for the technical support of T. Campbell, T. Davis, M. Foley and M. Vevea, and to all of our subjects. J.B.R. was supported by an HHMI predoctoral fellowship, and R.B.H.T. by the Human Frontiers program.

Correspondence and requests for materials should be addressed to J.B.R. at Harvard Medical School (e-mail: jreppas@nmr.mgh.harvard.edu).

The synaptic activation of kainate receptors

Michel Vignes & Graham L. Collingridge

Department of Anatomy, School of Medical Sciences, University of Bristol, University Walk, Bristol BS8 1TD, UK

L-Glutamate, the principal excitatory neurotransmitter in the vertebrate central nervous system, acts on three classes of ionotropic glutamate receptors, named after the agonists AMPA, NMDA and kainate¹. AMPA receptors are known to mediate fast synaptic responses and NMDA receptors to mediate slow synaptic responses at most excitatory synapses in the brain². Kainate receptors are formed from a separate set of genes (GluR5–7, KA-1 and KA-2) and are widely distributed throughout the brain^{3–8}. They are implicated in epileptogenesis and cell death⁹. However, the physiological functions of kainate receptors are not known⁷. The development of 2,3-benzodiazepine antagonists that are selective for AMPA receptors¹⁰ enables kainate receptors to be specifically activated by exogenous ligands, such as kainate^{11–16}. Here we demonstrate that high-frequency stimulation of mossy fibres in rat hippocampal slices, in the presence of the highly selective AMPA receptor antagonist GYKI 53655 (refs 13–15) plus NMDA- and GABA-receptor antagonists, activates an inward current in CA3 neurons that has a pharmacology typical of kainate receptors. The finding that kainate receptors can be activated synaptically adds to the diversity of information transfer at glutamatergic synapses.

Whole-cell recordings were obtained from CA3 neurons in response to single shock stimulation of the mossy fibre pathway and the associational/commissural pathway^{17,18} (Fig. 1a). In both pathways, AMPA-receptor-mediated excitatory postsynaptic currents (EPSCs) were blocked by the addition of GYKI 53655

(50 μ M), NMDA-receptor-mediated EPSCs were blocked by the selective antagonist AP5 (100 μ M) plus L-689,560 (5 μ M), GABA_A-receptor-mediated inhibitory postsynaptic currents (IPSCs) were blocked by picrotoxin (50 μ M) plus bicuculline (10 μ M), and GABA_B receptor-mediated IPSCs were eliminated by the use of Cs⁺-based intracellular solution. In the presence of this antagonist cocktail, essentially no synaptically evoked response remained in either pathway (Fig. 1b). However, high-frequency stimulation of the mossy fibre pathway (for example, 20 shocks at 100 Hz), using the same stimulus intensity, evoked a synaptic current, whereas equivalent stimulation of the associational/commissural input evoked little or no response (Fig. 1b). The size of the mossy fibre-evoked EPSC was dependent on the number of stimuli delivered. The synaptic response was clearly visible following paired stimuli and increased nonlinearly in response to between two and five stimuli. The maximum peak response was elicited by 10 shocks (Fig. 2a). The synaptic response had a linear *I*–*V* relationship and a reversal potential close to 0 mV ($+6 \pm 2$ mV; $n = 7$) (Fig. 2b).

GYKI 53655 is a potent AMPA-receptor antagonist, which, at a concentration of 50 μ M, does not significantly affect kainate receptors^{13–15}. The broad-spectrum ionotropic glutamate-receptor antagonist CNQX antagonizes kainate responses^{16,19}, and so can be used as a kainate antagonist under conditions where AMPA and NMDA receptors are already blocked. To test the effectiveness of CNQX as a kainate receptor antagonist on CA3 neurons, its ability to reverse kainate-induced currents was tested in the presence of the antagonist cocktail ($n = 8$, including 1 μ M tetrodotoxin in three experiments). Under these conditions, CNQX (10 μ M) rapidly and reversibly antagonized inward currents induced by kainate (200 nM) (Fig. 3a). Similarly, CNQX (10 μ M) also rapidly and reversibly antagonized the mossy fibre-evoked synaptic current ($n = 7$; Fig. 3b). At a concentration of 20 μ M, CNQX reduced the mossy fibre-evoked synaptic current by between 75 and 95% ($n = 7$; Fig. 3b, c).

It is unlikely that the mossy fibre-evoked response was a residual AMPA receptor-mediated response that appeared during high-frequency stimulation for the following reasons. First, GYKI 53655 was used at a concentration of approximately 50 times its half-maximal inhibitory concentration (IC₅₀) for antagonism at AMPA receptors^{13–15}, and is a non-competitive antagonist, so it should be independent of L-glutamate concentration. Second, the stimulus intensities were adjusted to evoke AMPA receptor-mediated EPSCs of a similar size in response to associational/commissural and mossy fibre stimulation, before the addition of the GYKI 53655-containing cocktail. A residual AMPA receptor-mediated EPSC should therefore have been as large after stimulation of the associational/commissural input. Third, CNQX was less potent and more rapidly reversible as an antagonist of the mossy fibre response than of an AMPA receptor-mediated EPSC recorded in brain slices. This difference correlates with the weaker potency of CNQX as a kainate as against an AMPA receptor antagonist^{16,19}. As a further test to distinguish between AMPA and kainate receptors we applied the AMPA receptor specific potentiator cyclothiazide²⁰. Before addition of the cocktail, cyclothiazide (3 μ M) produced the characteristic potentiation of a synaptic AMPA receptor-mediated EPSC²¹ in both the associational/commissural pathway (data not shown) and the mossy fibre pathway (Fig. 4a; $n = 5$). In contrast, when cyclothiazide was applied for the first time after addition of the cocktail, the mossy fibre-evoked EPSC was unaffected (Fig. 4b; $n = 4$). Activation of a residual NMDA receptor-mediated current can be excluded for several reasons. First, the mossy fibre-evoked synaptic current had a linear *I*–*V* relationship. Second, at the concentration used (10 μ M), CNQX does not antagonize NMDA receptor-mediated synaptic responses in hippocampal slices. Third, a much more potent glycine site antagonist than CNQX (L-689,560) was included in the cocktail.

It has been reported that stimulation of afferent fibres can evoke a

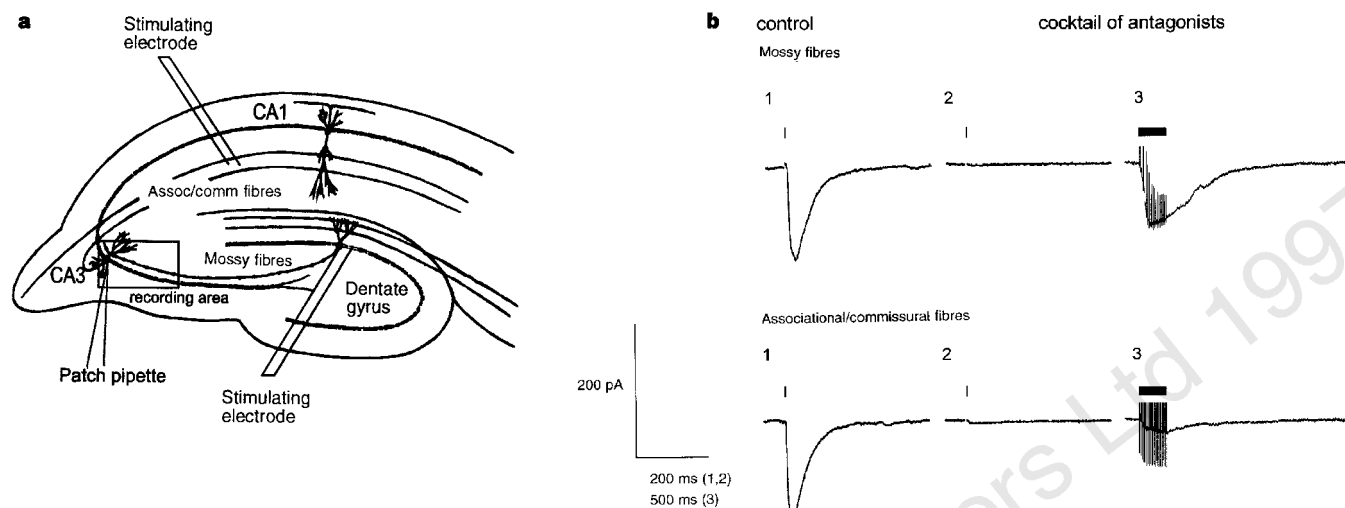


Figure 1 Synaptic activation of a non-AMPA, non-NMDA receptor-mediated EPSC by stimulation of mossy fibres. **a**, Schematic representation to show the positioning of the recording and stimulation electrodes. **b**, The traces show EPSCs recorded in control medium (1) and after addition of the antagonist cocktail (2 and 3), in response to low-frequency (1 and 2) and high-frequency (20 shocks, 100 Hz) (3) stimulation. In 12 neurons, single-shock stimulation, before addition of

the antagonist cocktail, evoked EPSCs of 173 ± 29 and 180 ± 12 pA in response to stimulation of the mossy fibre (top) and associational/commissural (bottom) pathways, respectively. After addition of the cocktail, high-frequency stimulation (10 or 20 shocks at 100 Hz, same intensity) evoked currents (measured 10 ms after the last stimulus) of 60 ± 9 and 13 ± 3 pA, respectively.

fast synaptic response in CA3 neurons that is mediated by activation of metabotropic glutamate receptors and is sensitive to (+)-MCPG²². However, the mossy fibre-evoked response observed in the present experiments was completely insensitive to (+)-MCPG ($500 \mu\text{M}$, $n = 4$; data not shown). It has also been shown that ATP can mediate fast EPSCs in the brain by means of a suramin-sensitive receptor²³. However, the mossy fibre-evoked EPSC was insensitive to antagonism by suramin ($30 \mu\text{M}$, $n = 4$; data not shown). It is possible that nicotinic receptors could mediate EPSCs in the brain, although the neuronal nicotinic receptor antagonist methyllycconitine (10 nM)²⁴ was also without effect on the mossy fibre-evoked EPSC ($n = 4$; data not shown).

Considered together, these data strongly suggest that the synaptic current is mediated through the activation of kainate receptors. The current was much larger after stimulation in the vicinity of the granule cells in the dentate gyrus, the cells of origin of the mossy fibres, than after stimulation within area CA1, which would activate mainly associational/commissural fibres. This finding correlates with the distribution of kainate receptors in the hippocampus²⁵. The residual current seen occasionally after associational/commissural stimulation may be due to inadvertent antidromic activation of some mossy fibres or to spill-over of L-glutamate from associational/commissural terminals to kainate receptors.

The frequency dependence of the mossy fibre-evoked EPSC is probably not due to the need for accumulation of glutamate in the synaptic cleft and diffusion to remote kainate receptors, because the glutamate uptake inhibitor L-trans-pyrrolidine-2,4-dicarboxylate ($100 \mu\text{M}$) had no effect on the mossy fibre-evoked EPSC ($n = 4$; data not shown). The frequency dependence could be due, at least in part, to the pronounced frequency facilitation exhibited by mossy fibres²⁶. An involvement of presynaptic facilitatory kainate autoreceptors¹⁶ is also a possibility.

Although highly enriched in the mossy fibre terminal zone of area CA3, kainate receptors are likely to be widely distributed throughout the nervous system^{8,25}. Thus many cell types respond to kainate under conditions where kainate rather than AMPA receptors can be assumed to mediate the response, because of the low dose of kainate used^{27,28} or the use of a 2,3-benzodiazepine to eliminate kainate actions on AMPA receptors^{11–16}. Furthermore, most neurons

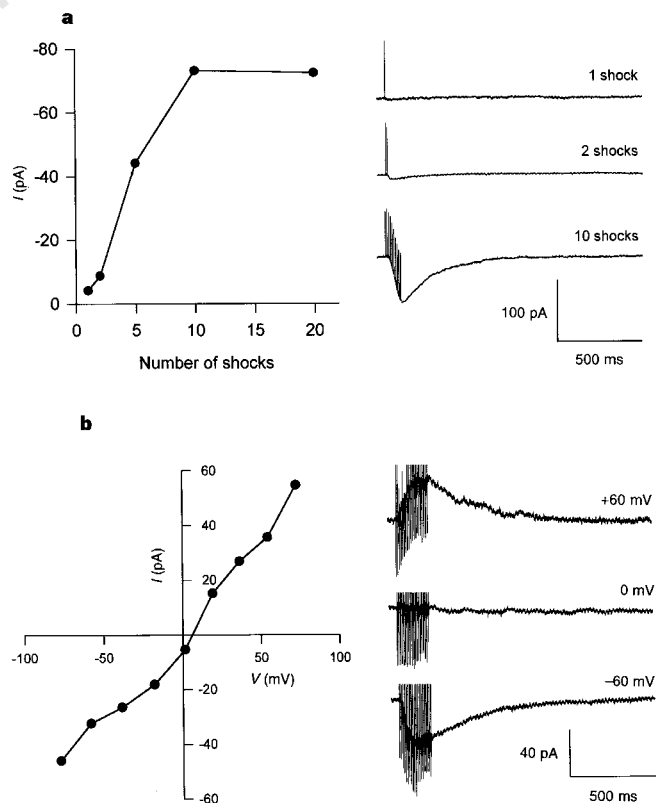


Figure 2 Properties of the mossy fibre-evoked EPSC. **a**, Peak amplitude of the EPSC versus number of stimuli used to elicit the response. The traces show mossy fibre-evoked by 1 (top), 2 (middle) and 10 (bottom) shocks delivered at 100 Hz. Similar results were seen in all 5 neurons tested. **b**, the mossy fibre-evoked EPSC has an approximately linear I/V relationship and a reversal potential close to 0 mV. The graph plots peak EPSC amplitude in response to 20 shocks delivered at 100 Hz versus holding potential.

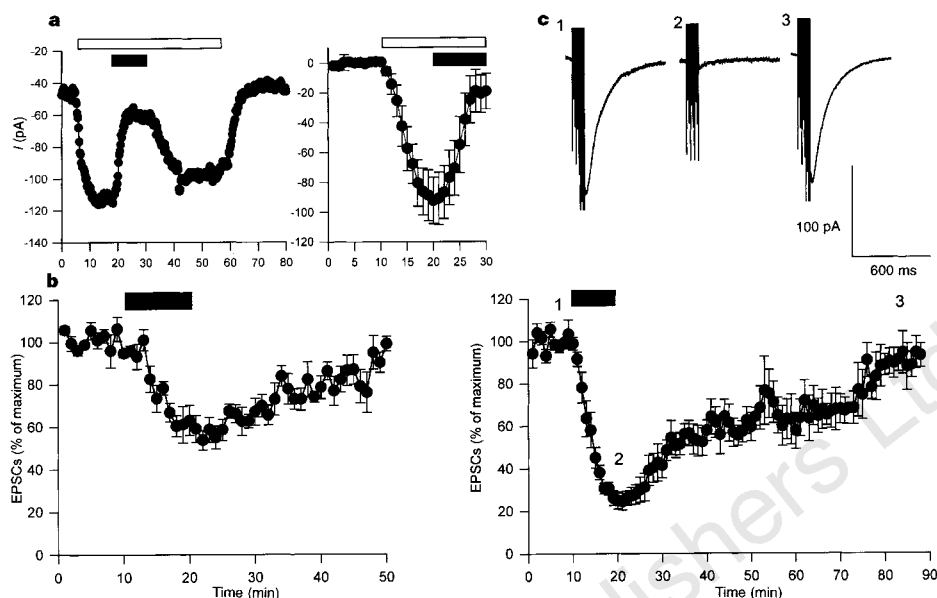


Figure 3 The mossy fibre EPSC is antagonized by CNQX. **a**, Kainate (200 nM; duration of application indicated by white bars) induces an inward current in the antagonist cocktail which is antagonized by CNQX (10 μ M; black bars). Left, a single example showing the reversibility of CNQX and kainate. Right, pooled data for 7 neurons. **b**, Mossy fibre EPSCs, evoked by 20 shocks delivered at 100 Hz, are

reversibly antagonized by CNQX (left, 10 μ M; right, 20 μ M). Each point plots the mean $\pm$ s.e. mean of the peak EPSC amplitude for 7 neurons. **c**, Traces of an experiment included in **b** (right) from a slice where CNQX (20 μ M) produced >90% antagonism of the synaptic response.

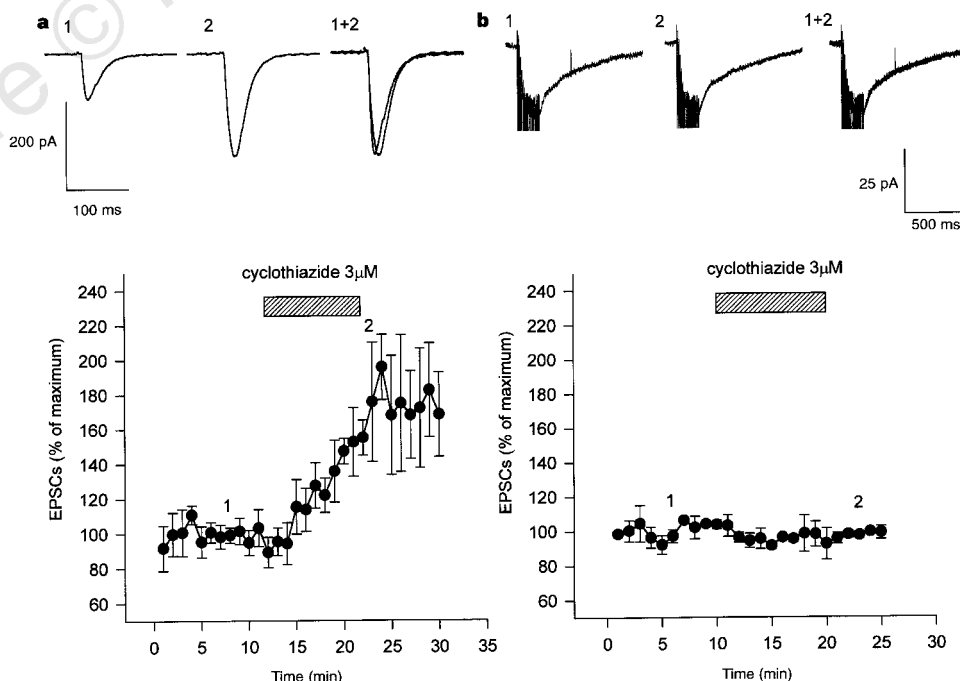


Figure 4 Cyclothiazide potentiates the mossy fibre-evoked EPSC in the absence, but not in the presence, of the antagonist cocktail. **a**, Responses to low-frequency stimulation of the mossy fibre pathway in control medium. Each point plots the mean $\pm$ s.e. mean of the peak EPSC amplitude for 5 neurons. Cyclothiazide (3 μ M) was applied for the duration of the bar and led to the characteristic

increase in response amplitude and an associated slowing of response kinetics. **b**, Responses to high-frequency stimulation (20 shocks at 100 Hz) of the mossy fibre pathway in the presence of the pharmacological cocktail. Each point plots the mean $\pm$ s.e. mean of the peak EPSC amplitude for 4 neurons.

express kainate receptor subunits⁸, and certain subunits have been localized on dendritic spines²⁹. It may therefore seem surprising that the synaptic activation of kainate receptors has not been reported previously. However, in most experiments the synaptic activation of AMPA receptors was eliminated using CNQX (or a related quinoxalinedione), and so kainate receptor-mediated responses would also have been antagonized.

The identity of the kainate receptor subtype(s) involved in mossy fibre synaptic transmission is not known, although it clearly involves a receptor subunit combination that does not exhibit rapid and complete desensitization and has a linear $I-V$ relationship^{6,12,30}. The function of kainate receptors at this synapse can only be a matter of speculation at present. However, the fact that their activation is facilitated during high-frequency transmission raises the possibility that kainate receptors are involved in synaptic plasticity in this pathway, which classically exhibits NMDA receptor-independent long-term potentiation^{17,18}. □

Methods

Experiments were performed on transverse hippocampal slices (400 μm) obtained by using a vibroslice¹⁶ from Wistar rats 13–17 days old. Slices were collected and maintained in medium comprising (mM): 124 NaCl, 3 KCl, 26 NaHCO₃, 1.25 NaH₂PO₄, 2 CaCl₂, 2 MgSO₄, 10 D-glucose (bubbled with O₂/CO₂: 95%/5%). After at least 1 h equilibration and recovery time, slices were transferred to a submerged style recording chamber and perfused with the same medium (at a rate of approximately 2 ml min⁻¹) at room temperature. Whole-cell patch-clamp recordings were obtained from the CA3 region, using the 'blind approach' with glass microelectrodes (5–7 M Ω ; seal resistance $\sim$ 10 G Ω) filled with a solution comprising (mM): 120 CsMeSO₄, 1 NaCl, 1 MgCl₂, 10 BAPTA, 5 N-(2,6-dimethyl-phenylcarbamoylmethyl)-trimethylammonium bromide (QX-314), 5 HEPES (adjusted to pH 7.3). In most experiments, 4 mM Mg-ATP was included, as it was found to prevent rundown of the synaptic currents. EPSCs were evoked by low-frequency stimulation of the mossy fibre and associational-commissural pathways alternately, using 20- μs pulses delivered through bipolar electrodes (Fig. 1a). Each pathway was stimulated once every 30 s, with 15 s separating the stimulation of each pathway. Two successive records per input were averaged to provide 1 min time points and for presentation of corresponding traces (the time points illustrated being indicated on the graphs by numbers). Apparent differences in stimulus artefacts are due to sampling. Access resistance was monitored continually and neurons discarded if this parameter changed by more than 20%. Unless otherwise stated, neurons were voltage-clamped at -60 mV. Drugs were applied by addition to the perfusing medium.

Received 2 December 1996; accepted 17 April 1997.

- Watkins, J. C. & Evans, R. H. Excitatory amino acid transmitters. *Annu. Rev. Pharmacol. Toxicol.* **21**, 165–204 (1981).
- Collingridge, G. L. & Lester, R. A. J. Excitatory amino acid receptors in the vertebrate central nervous system. *Pharmacol. Rev.* **40**, 143–210 (1989).
- Bettler, B. *et al.* Cloning of a novel glutamate receptor subunit, GluR5: Expression in the nervous system during development. *Neuron* **5**, 583–595 (1990).
- Egebjerg, J., Bettler, B., Hermans-Borgmeyer, I. & Heinemann, S. Cloning of a cDNA for a glutamate receptor subunit activated by kainate but not by AMPA. *Nature* **351**, 745–748 (1991).
- Werner, P., Voigt, M., Keinänen, K., Wisden, W. & Seeburg, P. H. Cloning of a high-affinity kainate receptor expressed predominantly in hippocampal CA3 cells. *Nature* **351**, 742–744 (1991).
- Herb, A. *et al.* The KA-2 subunit of excitatory amino acid receptors shows widespread expression in brain and forms ion channels with distantly related subunits. *Neuron* **8**, 775–785 (1992).
- Feldmeyer, D. & Cull-Candy, S. Neurotransmitters: Elusive glutamate receptors. *Curr. Biol.* **4**, 82–84 (1994).
- Bahn, S., Volk, B. & Wisden, W. Kainate receptor gene expression in the developing rat brain. *J. Neurosci.* **14**, 5525–5547 (1994).
- Meldrum, B. & Garthwaite, J. Excitatory amino acid neurotoxicity and neurodegenerative disease. *Trends Pharmacol. Sci.* **11**, 379–387 (1990).
- Tarnawa, I., Engberg, I. & Flatman, J. A. in *Amino acids: Chemistry, Biology and Medicine* (eds Lubec, G. & Rosenthal, G. A.) 538–546 (ESCOM Science Publishers, Leiden, 1990).
- Palmer, A. J. & Lodge, D. Cyclothiazide reverses AMPA receptor antagonism of the 2,3-benzodiazepine, GYKI-53655. *Eur. J. Pharmacol. Mol. Pharmacol. Section* **244**, 193–194 (1993).
- Renard, A., Crépel, F. & Audinat, E. Evidence for two types of non-NMDA receptors in rat cerebellar Purkinje cells maintained in slice cultures. *Neuropharmacology* **34**, 335–346 (1995).
- Paternain, A. V., Morales, M. & Lerma, J. Selective antagonism of AMPA receptors unmasks kainate receptor-mediated responses in hippocampal neurons. *Neuron* **14**, 185–189 (1995).
- Wilding, T. J. & Huettner, J. E. Differential antagonism of α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate-preferring and kainate-preferring receptors by 2,3 benzodiazepines. *Mol. Pharmacol.* **47**, 582–587 (1995).
- Bleakman, D. *et al.* Activity of 2,3-benzodiazepines at native rat and recombinant human glutamate receptors in vitro: Stereospecificity and selectivity. *Neuropharmacology* **35**, 1689–1702 (1996).

- Chittajallu, R. *et al.* Regulation of glutamate release by presynaptic kainate receptors in the hippocampus. *Nature* **379**, 78–81 (1996).
- Harris, E. W. & Cotman, C. W. Long-term potentiation of guinea pig mossy fibre responses is not blocked by N-methyl-D-aspartate antagonists. *Neurosci. Lett.* **70**, 132–137 (1986).
- Zalutsky, R. A. & Nicoll, R. A. Comparison of two forms of long-term potentiation in single hippocampal neurons. *Science* **248**, 1619–1624 (1990).
- Honoré, T. *et al.* Quinoxalinediones: potent competitive non-NMDA glutamate receptor antagonists. *Science* **241**, 701–703 (1988).
- Wong, L. A. & Mayer, M. L. Differential modulation of cyclothiazide and concanavalin-A of desensitization at native α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-preferring and kainate-preferring glutamate receptors. *Mol. Pharmacol.* **44**, 504–510 (1993).
- Rammes, G., Swandulla, D., Collingridge, G. L., Hartmann, S. & Parsons, C. G. Interactions of 2,3-benzodiazepines and cyclothiazide at AMPA receptors: patch clamp recordings in cultured neurones and area CA1 in hippocampal slices. *Br. J. Pharmacol.* **117**, 1209–1221 (1996).
- Pozzo-Miller, L. D., Petrozzino, J. J. & Connor, J. A. G protein-coupled receptors mediate a fast excitatory postsynaptic current in CA3 pyramidal neurons in hippocampal slices. *J. Neurosci.* **15**, 8320–8330 (1995).
- Edwards, F. A., Gibb, A. J. & Colquhoun, D. ATP receptor-mediated synaptic currents in the central nervous system. *Nature* **359**, 144–147 (1992).
- Gray, R., Rajan, A. S., Radcliffe, K. A., Yakehiro, M. & Dani, J. A. Hippocampal synaptic transmission enhanced by low concentrations of nicotine. *Nature* **383**, 713–716 (1996).
- Monaghan, D. T., Yao, D. & Cotman, C. W. L-[³H]glutamate binding to kainate-, NMDA-, and AMPA-sensitive sites: an autoradiographic analysis. *Brain Res.* **340**, 378–383 (1985).
- Salin, P. A., Scanziani, M., Malenka, R. C. & Nicoll, R. A. Distinct short-term plasticity at two excitatory synapses in the hippocampus. *Proc. Natl. Acad. Sci.* **93**, 13304–13309 (1996).
- Robinson, J. H. & Deadwyler, S. A. Kainic acid produces depolarization of CA3 pyramidal cells in the in vitro hippocampal slice. *Brain Res.* **221**, 117–127 (1981).
- Westbrook, G. L. & Lothman, E. W. Cellular and synaptic basis of kainic acid-induced hippocampal epileptiform activity. *Brain Res.* **273**, 97–109 (1983).
- Roche, K. W. & Huganir, R. L. Synaptic expression of the high-affinity kainate receptor subunit KA2 in hippocampal cultures. *Neuroscience* **69**, 383–393 (1995).
- Ruano, D., Lamboltz, B., Rossier, J., Paternain, A. V. & Lerma, J. Kainate receptor subunits expressed in single cultured hippocampal neurons: molecular and functional variants by RNA editing. *Neuron* **14**, 1009–1017 (1995).

Acknowledgements. We thank Eli Lilly for the GYKI 53655; W. W. Anderson for the on-line data acquisition and analysis programme, and J. C. Watkins for gifts of other compounds. This work was supported by the MRC and the European Community.

Correspondence and requests for materials should be addressed to G.L.C. (e-mail: g.l.collingridge@bristol.ac.uk).

Kainate receptors mediate a slow postsynaptic current in hippocampal CA3 neurons

Pablo E. Castillo^{*†}, Robert C. Malenka^{‡§} & Roger A. Nicoll^{*‡}

Departments of ^{*}Cellular and Molecular Pharmacology, [‡]Physiology and [§]Psychiatry, University of California at San Francisco, San Francisco, California 94143, USA

Glutamate, the neurotransmitter at most excitatory synapses in the brain, activates a variety of receptor subtypes that can broadly be divided into ionotropic (ligand-gated ion channels) and metabotropic (G-protein-coupled) receptors. Ionotropic receptors mediate fast excitatory synaptic transmission, and based on pharmacological and molecular biological studies are divided into NMDA and non-NMDA subtypes. The non-NMDA receptor group is further divided into AMPA and kainate subtypes¹. Virtually all fast excitatory postsynaptic currents studied so far in the central nervous system are mediated by the AMPA and NMDA subtypes of receptors. Surprisingly, despite extensive analysis of their structure, biophysical properties and anatomical distribution, a synaptic role for kainate receptors in the brain has not been found². Here we report that repetitive activation of the hippocampal mossy fibre pathway, which is associated with high-affinity kainate binding³ and many of the kainate receptor subtypes^{4–8}, generates a slow excitatory synaptic current with all of the properties expected of a kainate receptor. This activity-dependent synaptic current greatly augments the excitatory drive of CA3 pyramidal cells.

[†] Present address: Departamento de Fisiología, Facultad de Medicina, Gral Flores 2125, Montevideo, Uruguay.

We used standard electrophysiological techniques to determine whether the synaptic release of glutamate can activate kainate receptors. GYKI 53655 (1-(4-aminophenyl)-3-methylcarbonyl-4-methyl-7,8-methylenedioxy-3,4-dihydro-5H-2,3-benzodiazepine; 30 μ M), which is known to block AMPA (α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid)-receptor-mediated responses^{9,10} but not kainate receptor-mediated responses^{11,12}, blocks the mossy fibre field excitatory postsynaptic potential (EPSP) (>90%) (Fig. 1A,a). However, if repetitive stimuli are given in the presence of GYKI 53655, a slow component appears (Fig. 1A,b, trace 3), which is antagonized by CNQX (6-cyano-7-nitroquinoxaline-2,3-dione; 30 μ M) (Fig. 1A,b, trace 4). The results of this experiment are shown graphically in Fig. 1A,c. GYKI 53655 blocks the response to single stimuli, but in the continued presence of GYKI 53655, a response is restored by six stimuli at 30 Hz. This response is mediated by mossy fibres, as it is blocked by LCCG-1 (2S,3S,4S-CCG1(2S,1'S,2'S)-2-(carboxycyclopropyl)glycine; 10 μ M), which blocks mossy fibre responses by activating presynaptic inhibitory group II metabotropic glutamate receptors that are selectively expressed on mossy fibre terminals¹³. Finally, application of CNQX antagonizes the GYKI 53655-resistant slow component, confirming that it was mediated by non-NMDA receptors.

In this same experiment, responses to stimulation of the associational/commissural fibres were alternated with the mossy fibre responses (Fig. 1B). GYKI 53655 completely blocked the synaptic response to a single associational/commissural stimulus, leaving only a presynaptic fibre volley (Fig. 1B,a), and there was no recovery of a response with a repetitive stimulation (Fig. 1B,b). Thus CNQX had no effect in the presence of GYKI 53655. The finding that repetitive activation of mossy fibres, but not associational/commissural fibres elicits a GYKI 53655-resistant, CNQX-sensitive slow synaptic response was seen in eight separate experiments.

Although the mossy fibre slow component was blocked by CNQX (>80% at 30 μ M), it was about 10-fold less sensitive than the fast component (>80% at 3 μ M), consistent with previous data on kainate receptors^{5,14,15}. The slow component was also blocked by the non-NMDA receptor antagonists NBQX (10 μ M) ($n = 3$) and kynurenic acid (1 mM) ($n = 3$), but was unaltered by the NMDA-receptor antagonist AP5 (100 μ M) ($n = 5$), the metabotropic glutamate-receptor antagonist MCPG (1.5 mM) ($n = 4$), the glutamate transport blocker trans-PDC (500 μ M) ($n = 5$), or by NS-102 (3 μ M) ($n = 5$), an antagonist of low- but not high-affinity kainate receptors^{14,15}.

Whole-cell recording demonstrated that the presence and size of the GYKI 53655-resistant component were extremely sensitive to the frequency of stimulation. The GYKI 53655-resistant component was very small at low-frequency stimulation (Fig. 2A,a), but was always discernible with higher stimulus rates (Fig. 2A,b and 2A,c). When the stimulus frequency was changed from 0.05 Hz to 0.2 Hz, the response increased by $182 \pm 7\%$ ($n = 4$). This is very similar to the facilitation observed for the same change in frequency for the AMPA excitatory postsynaptic currents (EPSCs) ($194 \pm 13\%$, $n = 4$) (ref. 16). The decay of the GYKI 53655-resistant component ($\tau = 103 \pm 7$ ms, $n = 8$) was much slower than that for the GYKI 53655-sensitive component ($\tau = 9.1 \pm 1.2$ ms, $n = 6$) (Fig. 2A,d). In addition, the rise time (10–90%) of the GYKI 53655-resistant EPSC (6.8 ± 0.3 ms) was slower than the AMPA EPSC (2.9 ± 0.4 ms) recorded in the same cell ($n = 4$). The amplitude was strongly dependent both on the frequency of stimulation (Fig. 2B,a; $n = 6$) and on the number of stimuli in a train (Fig. 2B,b; $n = 6$). However, this enhancement was not associated with a change in the decay of the response. For example, τ at 1 Hz was 82 ms and at 5 Hz was 78 ms (Fig. 2B,a), and τ for two stimuli was 64 ms, and for three stimuli was 66 ms (Fig. 2B,b). GYKI 53655 completely blocked the EPSCs evoked by associational/commissural stimulation in the CA3 region (Fig. 2C; $n = 4$), as well as Schaffer collateral stimulation in the CA1 region (data not shown), and

repetitive stimulation did not restore transmission. The GYKI 53655-resistant component of the mossy fibre response was invariably blocked (>90%) by CNQX (50 μ M) (Fig. 2C; $n = 6$).

Further evidence consistent with the GYKI 53655-resistant slow EPSC being mediated by kainate receptors was provided by the action of cyclothiazide, which markedly potentiates AMPA receptor responses¹⁷ and AMPA receptor EPSCs in other regions¹⁸ yet has no effect on kainate receptor responses. Cyclothiazide enhanced the fast, GYKI 53655-sensitive EPSC (Fig. 2D,a; $n = 4$) but had no effect on the slow GYKI 53655-resistant EPSC (Fig. 2D,b; $n = 4$). Thus the response remaining in GYKI 53655 cannot be due to incomplete blockade of AMPA receptors. The GYKI 53655-resistant EPSC reverses at about 0 mV, and the I - V relationship exhibits slight outward rectification (Fig. 2E; $n = 4$). This finding suggests that the kainate receptors involved in this response are of the edited, Ca^{2+} -impermeable form, as unedited receptors show strong inward rectification¹⁹. The slow time course of the GYKI 53655-resistant EPSC raises the possibility that it might depend on the continued presence of glutamate. However, the time to half-decay of the EPSC was not increased by the glutamate uptake blocker trans-PDC (500 μ M) ($n = 4$), (average change = $+2.2 \pm 3.2\%$), which has been shown to enhance the spread of glutamate away from mossy fibre synapses²⁰.

We next tested for the presence of postsynaptic kainate receptors on CA3 pyramidal cells (Fig. 3A). CA3 pyramidal cells were approximately 30 times more sensitive to bath-applied kainate than were CA1 pyramidal cells, in agreement with previous results²¹. These high-affinity kainate receptors are localized at the mossy fibre

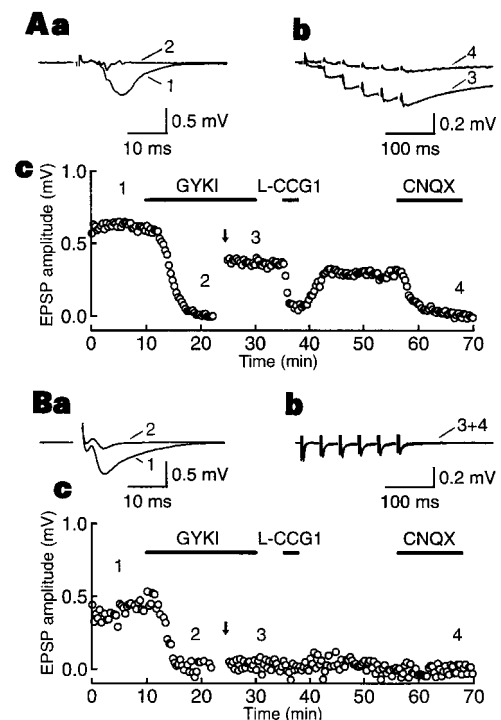


Figure 1 Repetitive activation of mossy fibres but not associational/commissural fibres activates a GYKI 53655 resistant, CNQX-sensitive synaptic potential, as revealed by field-potential recording. **A**, Mossy fibre field-potential records to single (**a**) and repetitive (**b**) stimulation were obtained at the times indicated by the numbers in the graph (**c**). Note the higher gain and slower time scale in **b** than in **a**. **B**, Associational/commissural responses were also recorded in this same experiment. Responses to single (**a**) and repetitive (**b**) stimulation were obtained at the times indicated by the numbers in the graph (**c**). The arrows in **A**, **c** and **B**, **c** indicate the time at which stimulation was changed from single stimuli (0.2 Hz) to repetitive stimuli (6 stimuli at 30 Hz given every 5 s). All traces are an average of 10–30 sweeps.

synapses (Fig. 3B), as brief puff application of kainate to stratum lucidum evoked a large inward current, whereas the identical application in stratum radiatum gave no response ($n = 4$). Repositioning the pipette in stratum lucidum again evoked an inward current. Puff application of kainate ($10 \mu\text{M}$) to stratum radiatum in the CA1 region close to the cell body layer failed to evoke any inward current ($n = 5$). Application of GYKI 53655, which completely blocked the AMPA receptor miniature EPSCs (downward deflections), had no effect on the stratum lucidum kainate response (Fig. 3C; $n = 4$). CNQX, however, completely blocked this response (Fig. 3C; $n = 4$). Like GYKI 53655, cyclothiazide also had no effect on the kainate responses (Fig. 3D,a), while having the expected enhancing effect on the size of the miniature EPSCs in the same experiment (Fig. 3D,b; $n = 3$). These results, using new pharmacological tools, directly demonstrate the presence of kainate receptors localized to the proximal dendrites of CA3 pyramidal cells, extending previous data on kainate responses associated with the mossy fibre projection²².

If the GYKI 53655-resistant EPSC is mediated by glutamate released onto kainate receptors, application of glutamate should also activate these receptors. Puff application of glutamate to stratum lucidum evoked a response that was typically only partially sensitive to GYKI 53655, even though the AMPA-receptor miniature EPSCs in this condition were entirely blocked (Fig. 4A). Addition of CNQX blocked the glutamate response remaining in GYKI (Fig. 4A; $n = 4$). In contrast, puff application of glutamate to stratum radiatum evoked an inward current that was completely blocked by GYKI 53655 (Fig. 4B; $n = 4$).

Although our results demonstrate the presence of postsynaptic kainate receptors on CA3 pyramidal cells, several studies have proposed that kainate receptors are present presynaptically not only on the mossy fibre terminals^{23–25}, but also on the terminals of

Schaffer collaterals in the CA1 region²⁶. Given the depolarizing action of these receptors, it can be anticipated that their activation on synaptic terminals would depolarize the terminals and as a consequence release glutamate, as has been demonstrated for presynaptic nicotinic receptors on mossy fibres²⁷. We therefore investigated whether application of kainate changes the frequency of AMPA-receptor miniature EPSCs, both when applied in stratum lucidum of the CA3 region and in stratum radiatum of the CA1 region. Puff application of kainate had little effect on the frequency of miniature EPSCs, either in the CA3 region (Fig. 4C,a) or in the CA1 region (Fig. 4C,b). Bath application of kainate also failed to alter the frequency of miniature EPSCs (Fig. 3A), whereas depolarization of the terminals with potassium (15 mM) increased the frequency considerably (data not shown). These results, together with the finding that GYKI 53655 and cyclothiazide had no effect on the size of the kainate-evoked response in CA3 cells, suggest that the predominant effect of kainate receptor activation is postsynaptic.

The finding that only the mossy fibres evoke a GYKI 53655-resistant EPSC, together with the selective action of kainate in stratum lucidum, provides a physiological basis for the remarkably selective distribution of high-affinity kainate binding to stratum lucidum³. If the kainate receptors in stratum lucidum have high affinity, why is it that at low-frequency stimulation most of the EPSCs are mediated by low-affinity AMPA receptors? If the AMPA and kainate receptors are colocalized at synaptic sites, the small size of the kainate receptor EPSC might be explained by a small single-channel conductance²⁸ and/or a low density of receptors, and the dramatic increase in the size of the response observed with high stimulus frequencies can be explained by the marked increase in transmitter release²⁰. Alternatively the kainate receptors could be excluded from the synaptic cleft, and with high stimulus frequencies

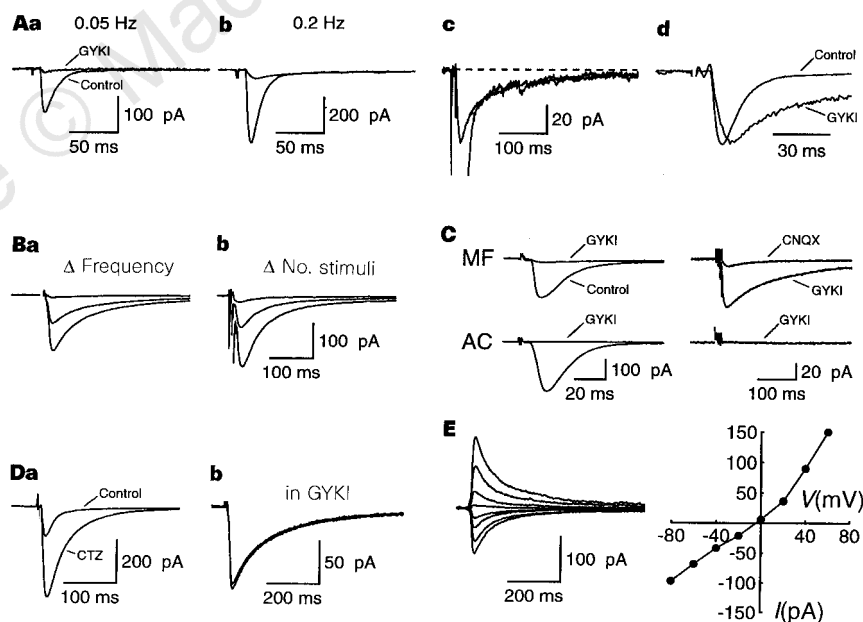


Figure 2 Characterization of the GYKI 53655 (GYKI)-resistant EPSC with whole-cell recording. **A**, Responses to two frequencies of stimulation in the absence and presence of GYKI ($30 \mu\text{M}$) are shown. All records are from the same cell. Increasing the frequency of stimulation from 0.05 Hz (**a**) to 0.2 Hz (**b**) greatly increases the size of the control and slow GYKI-resistant component. Note the decrease in gain for the records in **b**. **c**, Records in **b** at a much higher gain. **d**, The GYKI-resistant component has been scaled to the peak of the AMPA-receptor component. **B**, All responses were obtained in the presence of GYKI ($30 \mu\text{M}$). Increasing the frequency of stimulation (0.05 Hz , 1 Hz and 5 Hz) (**a**) or the number of stimuli in a train applied at 200 Hz (1 , 2 and 3 stimuli) (**b**) greatly increases the response. **C**, GYKI blocks the response to single mossy fibre (MF) and

associated/commissural (AC) stimulation (stimulus rate, 0.05 Hz) (left traces). Repetitive stimulation (4 stimuli at 200 Hz) (right traces) evokes a slow inward current at mossy fibres that is blocked by CNQX, but fails to restore a response at associational/commissural fibres. Note the increased gain and slower time scale for the records on the right. **D**, Cyclothiazide ($100 \mu\text{M}$) (CTZ) markedly increases the size of the mossy fibre AMPA-receptor EPSC (**a**) but has no effect on the mossy fibre EPSC evoked in the presence of GYKI (**b**). **E**, A family of superimposed current traces recorded in the presence of GYKI at different membrane potentials is shown. The peak current is plotted against membrane potential on the right. All traces are an average of 10 – 30 sweeps.

glutamate spills over onto these receptors. The similar degree of facilitation of the AMPA and kainate-mediated EPSCs, and the lack of change in the decay time constant of the kainate EPSC at high levels of glutamate release or when glutamate uptake is blocked²⁹, favours the first possibility. The slow kinetics of the GYKI 53655-resistant component contrasts with the rapid desensitization of kainate receptors, although functional receptors containing the KA1 subunit, which is selectively expressed in CA3 pyramidal cells⁴, have not been studied. Further investigation will be necessary to reconcile

the properties of kainate receptors to the kinetics of the EPSC mediated by these receptors.

Our results indicate that the synaptic release of glutamate from mossy fibres can activate postsynaptic kainate receptors as well as AMPA and NMDA receptors³⁰. The kainate-receptor EPSC shows a remarkable use-dependence, such that it is minimal at low-frequency stimulation. This use-dependence, as well as the slow kinetics of the kainate EPSC, provide a mechanism for augmenting pyramidal-cell excitability during high levels of granule-cell activity. It will be of

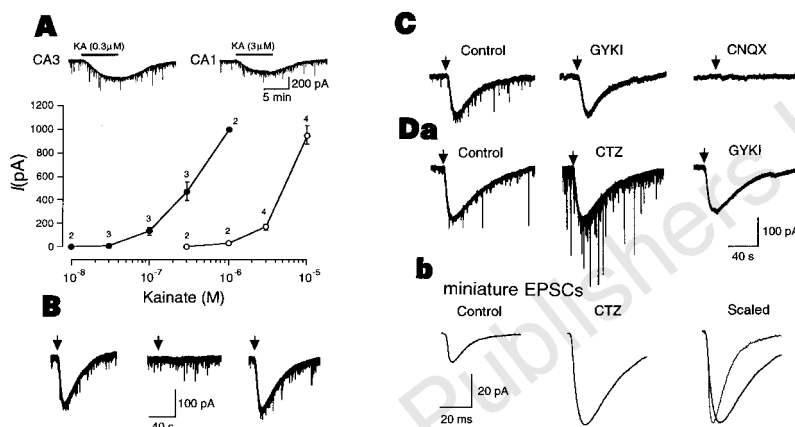


Figure 3 Characterization of responses to applied kainate. **A**, Graph illustrating the different sensitivity of CA1 (open circles) and CA3 (filled circles) pyramidal cells to bath-applied kainate. Responses to different concentrations of kainate were measured after the responses had reached equilibrium. The responses to the highest concentrations are underestimated because it was difficult to reach full equilibrium. All experiments were done in the presence of tetrodotoxin (0.3 μ M). Chart records of typical responses are shown above the graph. **B**, Puff application of kainate (3 μ M) on the surface of stratum lucidum (left) evokes an inward current. An identical puff applied 300 μ m away in stratum radiatum

(middle) evokes no response. Repositioning of the pipette over stratum lucidum (right) again evokes an inward current. **C**, Application of GYKI 53655 (GYKI) (30 μ M) has little effect on the response evoked by puff application of 3 μ M kainate but blocks the miniature EPSCs. The response is blocked by CNQX (25 μ M). **D**, Application of cyclothiazide (100 μ M) (CTZ) has no effect on the kainate response (**a**) but markedly increases the size of the miniature EPSCs seen as downward deflections (**a**) and as averaged traces (**b**). Miniature EPSCs were aligned and averaged (120 events each). Right, the control response has been scaled to the response obtained in cyclothiazide.

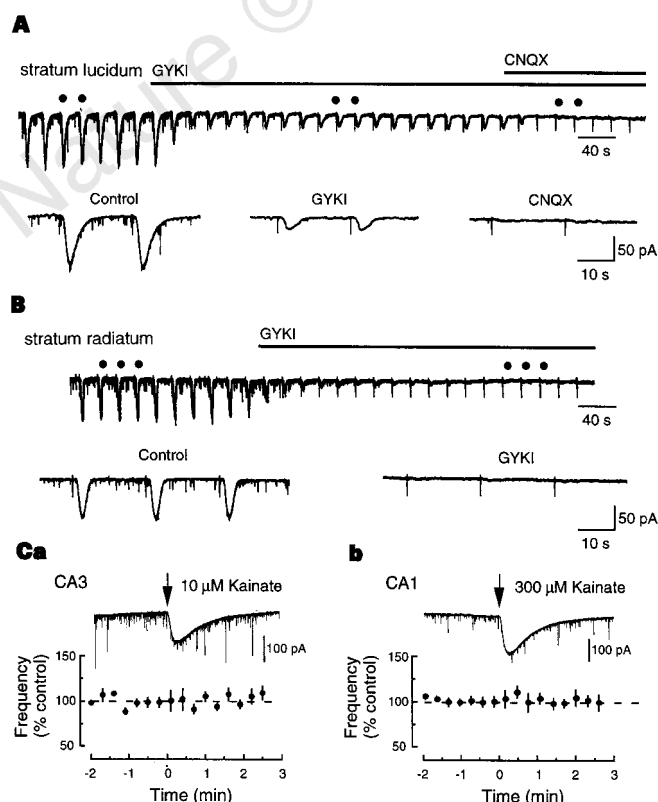


Figure 4 Characterization of the responses to glutamate and the lack of effect of kainate on the frequency of miniature EPSCs. **A**, Puff application of glutamate (3 mM) to the surface of stratum lucidum evokes an inward current that is only partly blocked by GYKI 53655 (GYKI) (30 μ M). The remaining current is blocked by CNQX (50 μ M). The responses identified by the filled circles above the continuous record are shown at a faster time scale beneath the continuous record. **B**, Puff application of glutamate (3 mM) to the surface of stratum radiatum evokes an inward current that is completely blocked by GYKI (30 μ M). Sample records at a faster time scale are shown below the continuous record. **C**, Puff application of 3 μ M kainate (200 kPa for 20 ms) to stratum lucidum in the CA3 region (**a**) or 300 μ M kainate to stratum radiatum in the CA1 region (**b**) fail to alter the frequency of miniature EPSCs. The sample traces are from a chart recorder, and are shown at the same speed as the graphs below. Both graphs plot the frequency of miniature EPSCs against time (4 cells for each graph).

interest to determine whether kainate receptors in other regions of the central nervous system, such as cerebellum and neocortex^{4-8,31}, have a similar function. □

Methods

Hippocampal slices (400–500 μm thick) were prepared from guinea-pigs (5–15 days old) and were maintained at room temperature for at least 1 h in a submerged chamber containing artificial cerebrospinal fluid (ACSF) equilibrated with 95% O_2 and 5% CO_2 . They were then transferred one at a time to a superfusing chamber³⁰. The ACSF contained (in mM): 119 NaCl, 2.5 KCl, 1 NaH_2PO_4 , 1.3 MgSO_4 , 2.5 CaCl_2 , 26 NaHCO_3 , 11 glucose. For all whole-cell recording experiments the MgSO_4 and CaCl_2 were increased to 4 mM and picrotoxin (100 μM) and AP5 (50 μM) were present. Experiments were done at room temperature. Whole-cell recording electrodes were filled with a solution containing (in mM): 122.5 caesium gluconate, 10 CsCl, 10 HEPES, 10 EGTA, 8 NaCl, 2 MgATP , 0.3 Na_3GTP , 1 CaCl_2 , 10 glucose. The resistance of the patch pipette ranged between 2 and 4 $\text{M}\Omega$. Access resistances ranged between 8 and 15 $\text{M}\Omega$, and were not allowed to vary by more than 15% during the course of the experiment. No series resistance compensation was used. Field recordings were performed with a glass electrode placed in the stratum lucidum. Bipolar tungsten electrodes were placed in the granule-cell layer to stimulate mossy fibres and in the stratum radiatum to stimulate associational/commissural fibres. A picospritzer was used to puff-apply glutamate and kainate. The pressure varied from 100 to 200 kPa, and the duration of the puff varied from 20 to 100 ms. The drugs were prepared in Fast Green (0.01%) (Sigma) to help localize the application visually. Average values are expressed as mean $\pm$ s.e.m. Drugs used were: L-glutamic acid, kainate, kynurenate, cyclothiazide, picrotoxin, L-AP4 (Sigma), CNQX, NBQX, D-AP5, (+)-MCPG and LCCG-1 (Tocris Cookson), L-trans-pyrrolidine-2,4-dicarboxylic acid, NS-102 (Research Biochemicals International), and tetrodotoxin (CalBiochem).

Received 19 November 1996; accepted 22 April 1997.

- Bettler, B. & Mülle, C. Review: Neurotransmitter receptors II. AMPA and kainate receptors. *Neuropharmacology* **34**, 123–139 (1995).
- Lerma, J., Morales, M., Vicente, M. A. & Herreras, O. Glutamate receptors of the kainate type and synaptic transmission. *Trends Neurosci.* **20**, 9–12 (1997).
- Foster, A. C., Mena, E. E., Monaghan, D. T. & Cotman, C. W. Synaptic localization of kainic acid binding sites. *Nature* **289**, 73–75 (1981).
- Werner, P., Voigt, M., Keinänen, K., Wisden, W. & Seeburg, P. H. Cloning of a putative high-affinity kainate receptor expressed predominantly in hippocampal CA3 cells. *Nature* **351**, 742–744 (1991).
- Egebjerg, J., Bettler, B., Hermans-Borgmeyer, I. & Heinemann, S. Cloning of a cDNA for a glutamate receptor subunit activated by kainate but not AMPA. *Nature* **351**, 745–748 (1991).
- Herb, A. et al. The KA-2 subunit of excitatory amino acid receptors shows widespread expression in brain and forms ion channels with distantly related subunits. *Neuron* **8**, 775–785 (1992).
- Petralia, R. S., Wang, Y.-X. & Wenthold, R. J. Histological and ultrastructural localization of the kainate receptor subunits, KA2 and GluR6/7, in the rat nervous system using selective antipeptide antibodies. *J. Comp. Neurol.* **349**, 85–110 (1994).
- Siegel, S. J. et al. Distribution of the excitatory amino acid receptor subunits GluR2(4) in monkey hippocampus and colocalization with subunits GluR5-7 and NMDAR1. *J. Neurosci.* **15**, 2707–2719 (1995).
- Donevan, S. D. & Rogawski, M. A. GYKI 52466, a 2,3-benzodiazepine, is a highly selective, noncompetitive antagonist of AMPA/kainate receptor responses. *Neuron* **10**, 51–59 (1993).
- Zorumski, C. F., Yamada, K. A., Price, M. T. & Olney, J. W. A benzodiazepine recognition site associated with the non-NMDA glutamate receptor. *Neuron* **10**, 61–67 (1993).
- Paternalin, A. V., Morales, M. & Lerma, J. Selective antagonism of AMPA receptors unmasks kainate receptor-mediated responses in hippocampal neurons. *Neuron* **14**, 185–189 (1995).
- Wilding, T. J. & Huettner, J. E. Differential antagonism of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-preferring and kainate-preferring receptors by 2,3-benzodiazepines. *Mol. Pharmacol.* **47**, 582–587 (1995).
- Kamiya, H., Shinokaki, H. & Yamamoto, C. Activation of metabotropic glutamate receptor type 2/3 suppresses transmission at rat hippocampal mossy fibre synapses. *J. Physiol. (Lond.)* **493**, 447–455 (1996).
- Verdoorn, T. A., Johansen, T. H., Drejer, J. & Nielsen, E. O. Selective block of recombinant glutamate receptors by NS-102, a novel non-NMDA receptor antagonist. *Eur. J. Pharmacol.* **269**, 43–49 (1994).
- Lerma, J., Paternalin, A. V., Narajio, J. R. & Mellstrom, B. Functional kainate-selective glutamate receptors in cultured hippocampal neurons. *Proc. Natl Acad. Sci. USA* **90**, 11688–11692 (1993).
- Salin, P. A., Scanziani, M., Malenka, R. C. & Nicoll, R. A. Distinct short-term plasticity at two excitatory synapses in the hippocampus. *Proc. Natl Acad. Sci. USA* **93**, 13304–13309 (1996).
- Partin, K. M., Patneau, D. K., Winters, C. A., Mayer, M. L. & Buonanno, A. Selective modulation of desensitization at AMPA versus kainate receptors by cyclothiazide and concanavalin A. *Neuron* **11**, 1069–1082 (1993).
- Trussell, L. O., Zhang, S. & Raman, I. M. Desensitization of AMPA receptors upon multiquantal neurotransmitter release. *Neuron* **10**, 1185–1196 (1993).
- Burnashev, N., Villarroel, A. & Sakmann, B. Dimensions and ion selectivity of recombinant AMPA and kainate receptor channels and their dependence on Q/R site residues. *J. Physiol. (Lond.)* **496**, 165–173 (1996).
- Scanziani, J. H. & Salin, P. A., Vogt, K. E., Malenka, R. C. & Nicoll, R. A. Use-dependent increases in glutamate concentration activate presynaptic metabotropic glutamate receptors. *Nature* **385**, 630–634 (1997).
- Robinson, J. H. & Deadwyler, S. A. Kainic acid produces depolarisation of CA3 pyramidal cells in the in vitro hippocampal slice. *Brain Res.* **221**, 117–127 (1981).

- Sawada, S., Higashima, M. & Yamamoto, C. Kainic acid induces long-lasting depolarizations in hippocampal neurons only when applied to stratum lucidum. *Exp. Brain Res.* **72**, 135–140 (1988).
- Sawada, S., Higashima, M. & Yamamoto, C. Inhibitors of high-affinity uptake augment depolarizations of hippocampal neurons induced by glutamate, kainate, and related compounds. *Exp. Brain Res.* **60**, 323–329 (1985).
- Reprea, A., Tremblay, E. & Ben-Ari, Y. Kainate binding sites in the hippocampal mossy fibers: Localization and plasticity. *Neuroscience* **20**, 739–748 (1987).
- Malva, J. O. et al. A functionally active presynaptic high-affinity kainate receptor in the rat hippocampal CA3 subregion. *Neurosci. Lett.* **185**, 83–86 (1995).
- Chittajallu, R. et al. Regulation of glutamate release by presynaptic kainate receptors in the hippocampus. *Nature* **379**, 78–81 (1996).
- Gray, R., Rajan, A. S., Radcliffe, K. A., Yakehiro, M. & Dani, J. A. Hippocampal synaptic transmission enhanced by low concentrations of nicotine. *Nature* **383**, 713–716 (1996).
- Swanson, G. T., Feldmeyer, D., Kaneda, M. & Cull-Candy, S. G. Effect of RNA editing and subunit co-assembly on single-channel properties of recombinant kainate receptors. *J. Physiol. (Lond.)* **492**, 129–142 (1996).
- Otis, T. S., Wu, Y.-C. & Trussell, L. O. Delayed clearance of transmitter and the role of glutamate transporters at synapses with multiple release sites. *J. Neurosci.* **16**, 1634–1644 (1996).
- Weisskopf, M. G. & Nicoll, R. A. Presynaptic changes during mossy fibre LTP revealed by NMDA receptor-mediated synaptic responses. *Nature* **376**, 256–259 (1995).
- Huntley, G. W. et al. Selective distribution of kainate receptor subunit immunoreactivity in monkey neocortex revealed by a monoclonal antibody that recognizes glutamate receptor subunits GluR5/6/7. *J. Neurosci.* **13**, 2965–2981 (1993).

Acknowledgements. We thank D. Leander for GYKI 53655; M. Frerking, S. Olet and K. Vogt for comments on the manuscript; and H. Czerwonka for secretarial assistance. R.A.N. is a member of the Keck Center for Integrative Neuroscience and the Silvio Conte Center for Neuroscience Research. R.C.M. is a member of the Center for Neurobiology and Psychiatry, and the Center for the Neurobiology of Addiction. R.A.N. and R.C.M. are supported by grants from the NIH.

Correspondence and requests for materials should be addressed to R.A.N. (e-mail: nicoll@phy.ucsf.edu).

Muscle force is generated by myosin heads stereospecifically attached to actin

Sergey Y. Bershitsky*, Andrey K. Tsaturyan†, Olga N. Bershitskaya*, Gregory I. Mashanov*, Paul Brown, Ronald Burns & Michael A. Ferenczi

National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

* Institute of Physiology, Urals Branch of the Russian Academy of Sciences, Yekaterinburg 620219, Russia

† Institute of Mechanics, Lomonosov Moscow University, Moscow 119899, Russia

Muscle force is generated by myosin crossbridges interacting with actin. As estimated from stiffness^{1,2} and equatorial X-ray diffraction³ of muscle and muscle fibres, most myosin crossbridges are attached to actin during isometric contraction, but a much smaller fraction is bound stereospecifically⁴⁻⁷. To determine the fraction of crossbridges contributing to tension and the structural changes that attached crossbridges undergo when generating force, we monitored the X-ray diffraction pattern during temperature-induced tension rise in fully activated permeabilized frog muscle fibres. Temperature jumps⁸ from 5–6 °C to 16–19 °C initiated a 1.7-fold increase in tension without significantly changing fibre stiffness or the intensities of the (1,1) equatorial and (14.5 nm)⁻¹ meridional X-ray reflections. However, tension rise was accompanied by a 20% decrease in the intensity of the (1,0) equatorial reflection and an increase in the intensity of the first actin layer line by ~13% of that in rigor. Our results show that muscle force is associated with a transition of the crossbridges from a state in which they are nonspecifically attached to actin to one in which stereospecifically bound myosin crossbridges label the actin helix.

Muscle force results from changes in myosin crossbridges, or 'heads', that are attached to thin filaments^{9,10}. Changes in the intensities, $I_{1,0}$, $I_{1,1}$, of the main equatorial X-ray reflections³ (1,0) and (1,1), and stiffness measurements^{1,2} show that 50–75% of myosin heads are attached to the thin filaments in fully activated muscle contracting isometrically. Surprisingly, the intensity, I_{1a} , of

the first actin layer line at about $(37 \text{ nm})^{-1}$, which arises from myosin heads labelling the actin helix, is only 3–10% of that in rigor^{4–7,11}. As I_{1a} is proportional to the square of the number of diffractors, this suggests that only 17–32% of myosin heads bind to actin stereospecifically. Electron microscopy¹², fluorescence polarization¹³ and electron paramagnetic resonance (EPR) spectroscopy¹⁴ also show a small fraction of myosin crossbridges bound to actin in a rigor-like manner in contracting muscle. It has been suggested, but never directly confirmed, that only specifically bound crossbridges participate in force generation¹¹.

To observe the structural changes in crossbridges during force generation, it is necessary to synchronize the power strokes in a large fraction of the crossbridges. At the onset of contraction, force generation results from activation of thin filaments and an increase in the number of attached crossbridges¹⁵, making it difficult to distinguish the structural changes associated with each process. Another approach is the application of sudden length changes to contracting muscle^{10,16}. A significant decrease in the intensity of the $(14.5 \text{ nm})^{-1}$ meridional reflection, $I_{14.5}$, but not in the actin-based layer lines or equatorial X-ray reflections, followed the length steps¹⁷. In single muscle fibres^{18–20}, it was found that the decrease in $I_{14.5}$ had the same time course as that of the phase of fast-tension recovery after the step. As the $(14.5 \text{ nm})^{-1}$ reflection arises from axial repeat of the heads along the thick filament, these results support the idea that fast length perturbations induce a tilt in a majority of attached crossbridges. If such a tilt is the process underlying isometric force generation^{10,18}, then the large change in $I_{14.5}$ after a length perturbation indicates that this signal is caused by a large fraction of the muscle crossbridges, an interpretation that is not compatible with a contribution to isometric tension by only a small fraction of the crossbridges stereospecifically bound to actin.

To determine the change in crossbridges associated with force generation, we changed the temperature rapidly (temperature jump)⁸ to increase tension without perturbing the relative positions of the thick and thin filaments. Structural changes were followed by time-resolved low-angle X-ray diffraction using synchrotron radiation. Contracting permeabilized muscle fibres of the frog are prone to rapid loss of structure. A procedure involving partial crosslinking of the myofilaments with EDC (see Methods) was used²¹ to stabilize the preparation while maintaining its mechanical and physiological properties. To compare the changes induced by the two synchronization methods, length steps were imposed before and after temperature jump (Fig. 1).

Contracting fibres were released in $140 \mu\text{s}$ by $4 \pm 1.1 \text{ nm}$ (mean $\pm$ s.d., $n = 13$, unless otherwise specified) per half sarco-

mere and re-stretched to their initial length after 52 ms (Fig. 1). The rate constant of the fast partial tension recovery following the release was $1,450 \pm 245 \text{ s}^{-1}$ at $\sim 6^\circ\text{C}$. After the stretch it was $578 \pm 95 \text{ s}^{-1}$. Both releases and stretches induced a transitory decrease in $I_{14.5}$, as found for intact muscles¹⁷ and muscle fibres¹⁹. The amplitude of the decrease in $I_{14.5}$ was close to that found in intact fibres²⁰. $I_{1,0}$ and $I_{1,1}$ were slightly changed by the length steps (Fig. 1). When tension and $I_{14.5}$ had recovered to almost isometric levels, the fibre temperature was increased from $5\text{--}6^\circ\text{C}$ to $16\text{--}19^\circ\text{C}$ in 0.35 ms and then maintained at the higher temperature (Fig. 1). This increased tension by a factor of 1.68 ± 0.16 without significantly changing sarcomere length in the part of the fibre exposed to the X-ray (Fig. 1). Fibre stiffness measured from the changes in tension and sarcomere length during stretches complete in $140 \mu\text{s}$ (Fig. 1), or from 2.5-kHz sinusoidal oscillations (0.05% peak-to-peak of the fibre length), changed by $\leq 3\%$ after the temperature jump. The tension transient was fitted by the sum of two exponentials with rate constants of $1,800 \pm 460 \text{ s}^{-1}$ and $121 \pm 30 \text{ s}^{-1}$. The amplitude of the fast component was $53.8 \pm 4.4\%$ of the tension increase.

At the plateau of the isometric tension at the higher temperature, $I_{14.5}$ and $I_{1,1}$ were 1.06 ± 0.1 and 1.04 ± 0.03 of the intensities at the lower temperature, respectively (Figs 1, 2). The widths of the $(1,0)$, $(1,1)$ equatorial and of the $(14.5 \text{ nm})^{-1}$ meridional reflections changed by $\leq 10\%$ in both meridional and equatorial directions, showing that the temperature-induced change in disorder of the second kind²² did not affect their intensities. The temperature jump shrank the filament lattice by $0.5\text{--}1\%$, as estimated from the separation of the $(1,0)$ and $(1,1)$ equatorial reflections.

The temperature jump induced a $20 \pm 6\%$ decrease in $I_{1,0}$ (Fig. 2), with an apparent rate constant of $340 \pm 10 \text{ s}^{-1}$ which overall was faster than the increase in tension. However, the time resolution was insufficient to determine whether the change in $I_{1,0}$ was simultaneous with the fast component of tension rise.

The release–stretch cycle was repeated in the higher temperature when tension and the X-ray reflections had reached a new steady state. The rate constants of the fast tension recovery following the release and stretch were $2,860 \pm 720 \text{ s}^{-1}$ and $1,540 \pm 200 \text{ s}^{-1}$, respectively. The changes in $I_{14.5}$ induced by the length perturbations were also faster at the higher temperature; again, the changes in $I_{1,0}$ and $I_{1,1}$ were small (Fig. 1).

The first actin layer line at about $(37 \text{ nm})^{-1}$ was much weaker than in rigor and not resolved from the remainder of the $(43 \text{ nm})^{-1}$ myosin layer line (Fig. 4a), as described for intact muscle^{4,5}. However, the difference in the off-meridional intensities, ΔI_{1a} , of

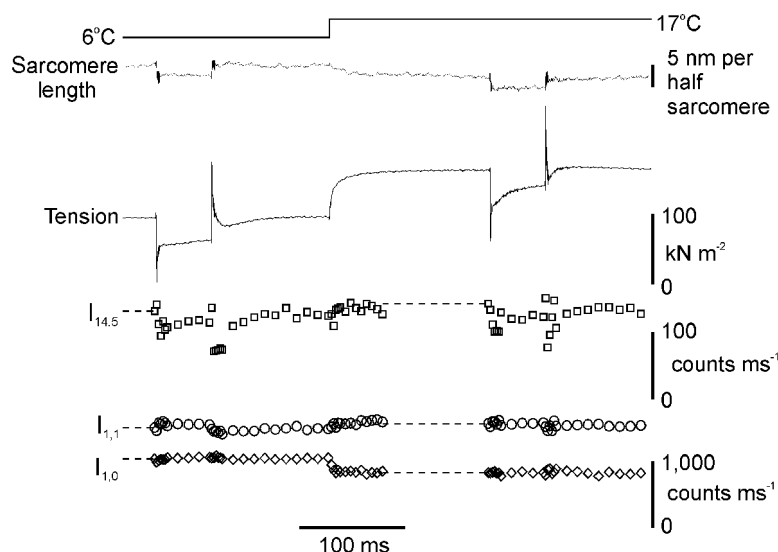


Figure 1 Time courses (traces from top to bottom) of temperature, change in sarcomere length, tension and intensities of the X-ray reflections $I_{14.5}$ ($\square$), $I_{1,0}$ ($\diamond$) and $I_{1,1}$ ($\circ$) during the length-steps and temperature jump (T-jump) protocol. Data points are shown at 2, 5 or 10-ms intervals; dashed lines show X-ray intensities during two 100-ms time frames recorded under isometric conditions at low and high temperature. Temperature, tension and sarcomere length records are from a representative experiment (fibre dimensions, $3.67 \text{ mm} \times 16,040 \mu\text{m}^2$; sarcomere length, $2.1 \mu\text{m}$). The slow sarcomere shortening following the T-jump probably results from heat loss through the fibre ends, resulting in some shortening of the warmer central region of the fibre.

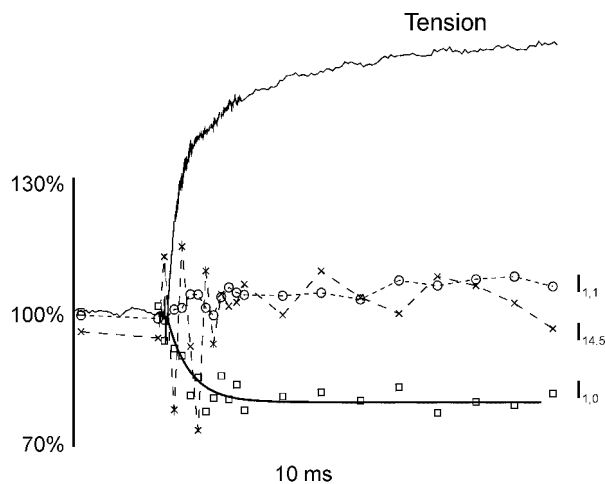


Figure 2 Normalized tension (continuous line), meridional $I_{14.5}$ (crosses), equatorial $I_{1.0}$ (squares) and $I_{1.1}$ (circles) intensities following the T-jump. Tension and the intensities of the X-ray reflections were normalized to their level in the 100-ms frame during the isometric contraction at 5–6°C (Fig. 1). The thick line is an exponential fit of the time course of $I_{1.0}$, giving a rate constant of $340 \pm 110 \text{ s}^{-1}$. The X-ray data were taken at 1-ms intervals.

the X-ray diffraction patterns at the plateau of isometric contraction at 16–19°C and 5–6°C (Fig. 3) shows a peak at about $(37 \text{ nm})^{-1}$. The position of the centre of the peak coincided with that of the first actin layer line in rigor and was distinct from the position of the first myosin layer line. Thus, the temperature jump intensified the first actin layer line. From the gaussian fit (Fig. 4b), the increase in intensity ΔI_{1a} was $13.5 \pm 3.4\%$ (mean $\pm$ s.e.m.; Fig. 4b) of I_{1a} in rigor, in which all myosin heads bind strongly to actin²³. The intensity of the background surrounding the layer lines in the difference diffraction pattern is negative (Fig. 4b), suggesting that crossbridges are less disordered at the higher temperature, where they generate more force.

The increase in I_{1a} following the temperature jump shows that force generation is accompanied by an increase in the fraction of myosin heads that label the actin helix. As stiffness and $I_{1.1}$ do not change, the I_{1a} increase is not due to new attachment of crossbridges, but is a transition of attached crossbridges to a stereospecific state. Assuming that force is proportional to the fraction of stereospecifically bound crossbridges and that they alone contribute to the first actin layer line, then their fraction, x , before the temperature jump can be estimated as follows. The jump causes an increase in the fraction of stereospecifically bound crossbridges from x to $1.68x$. I_{1a} is proportional to x^2 and reaches its maximum in rigor, when all crossbridges label the actin helix. So before the temperature jump, the fraction of stereospecifically bound heads is given by $(1.68x)^2 - x^2 = \Delta I_{1a} = 0.135$ and $x = 0.27$, which increases to $1.68x = 0.46$ after the jump.

If as many as 20% of the myosin heads are crosslinked by

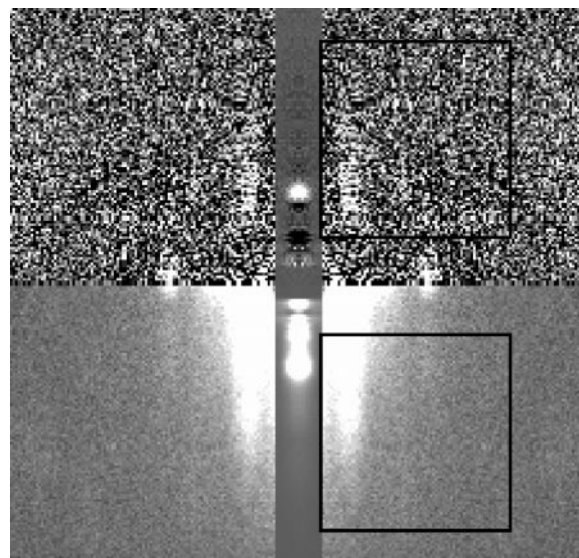


Figure 3 Two-dimensional diffraction patterns. Top, difference between X-ray diffraction patterns obtained at high (16–19°C) and low (5–6°C) temperature at the plateau of isometric tension in 100-ms frames (Fig. 1) before and after the T-jump. Bottom, rigor pattern attenuated by a factor of 10. White indicates higher intensity; the patterns are symmetrically averaged. The equator is vertical, and attenuated in software by a factor of 20 for clarity. Black rectangles show the I_{1a} integration area used for Fig. 4, corresponding to radial reciprocal spacing of $(35 \text{ nm})^{-1}$ to $(8 \text{ nm})^{-1}$ and a meridional region from $(70 \text{ nm})^{-1}$ to $(9 \text{ nm})^{-1}$. The T-jump-induced increase in I_{1a} appears as a white vertical stripe in the rectangle of the upper half. The decrease in $I_{1.0}$ at the higher temperature is seen as black equatorial spots near the centre of the pattern. Data for active contractions were obtained from 6 muscle fibres, with a total exposure of 21.8 s at each of the two temperatures. The rigor pattern was obtained at the start of each experiment from the same 6 fibres, total exposure of 12 s, by summing the pattern from each fibre after normalization for their exposure.

EDC in a rigor-like configuration, and if all these label the actin helix and do not contribute to tension²¹, this equation becomes $(0.2 + 1.68x')^2 - (0.2 + x')^2 = \Delta I_{1a} = 0.135$, where x' is the fraction of specifically bound non-crosslinked crossbridges before the temperature jump. The solution is $x' = 0.21$ or $0.21/0.8 = 26\%$ of the number of unlinked crossbridges: the result is independent of any contribution of crosslinked heads to the intensity of the first actin layer line.

The decrease in $I_{1.0}$ that accompanies the tension rise induced by the temperature jump (Figs 1, 2) also indicates that force generation is associated with a radial and/or azimuthal movement of crossbridges. It was suggested^{24,25} that $I_{1.0}$, but not $I_{1.1}$ is sensitive to a transition of the crossbridges from weak, non-stereospecific attachment to the thin filaments to strong, stereospecific binding. This transition moves the centre of mass of the crossbridges away from the backbone of the thick filaments and closer to the thin filaments, decreasing $I_{1.0}$. The electron density of the (1,1) planes passing through the thin and thick filaments remains nearly constant and $I_{1.1}$ does not change.

The structural changes underlying tension transients following temperature or length perturbation are not equivalent. The length steps induce a significant decrease in $I_{14.5}$, but not in I_{1a} , $I_{1.0}$ or in $I_{1.1}$ (refs 17, 18) (Figs 1, 2). In contrast, the temperature jump does not affect $I_{14.5}$ but initiates a decrease in $I_{1.0}$ and an increase in I_{1a} . An interpretation of the decrease in $I_{14.5}$ after stretches or releases¹⁸ is that a large fraction of the myosin crossbridges tilt to become more aligned with the fibre axis. Fluorescent probes attached to myosin light chains also tilt during and following length changes²⁶. Is this

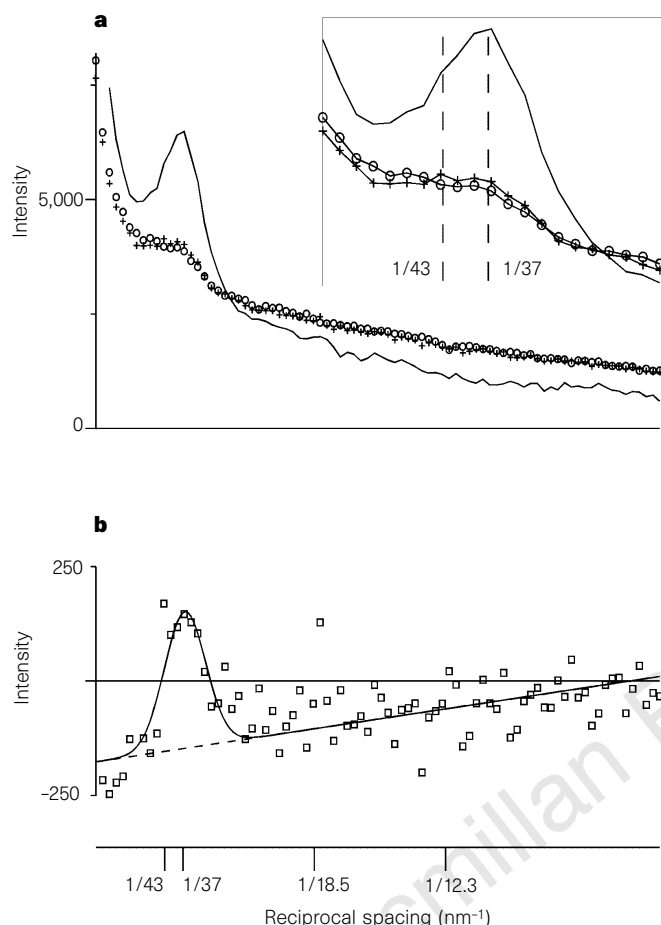


Figure 4 a, Off-meridional X-ray intensities (counts per pixel) integrated along the equator between $(35 \text{ nm})^{-1}$ and $(8 \text{ nm})^{-1}$ and plotted against the meridional spacing. Data were obtained from the diffraction patterns used in Fig. 3 during active contraction at the higher (+) and lower (O) temperatures and in rigor (continuous line). Inset, region around the first actin layer line on an extended scale. Vertical dashed lines correspond to the reciprocal spacing of the first myosin $(43 \text{ nm})^{-1}$ and actin $(37 \text{ nm})^{-1}$ layer lines. The intensity of I_{1a} in rigor after linear background subtraction was fitted by a gaussian centred at $(37.0 \pm 0.1 \text{ nm})^{-1}$ with amplitude $2,578 \pm 37$ and half-width $(3.56 \pm 0.06) \times 10^{-3} \text{ nm}^{-1}$. **b**, Difference between the off-meridional intensities at the higher and lower temperature shown in **a** ($\square$). The intensity is scaled by 10 compared to **a**. The continuous line is a least-square fit with a gaussian peak and a linear background (dashed line). The gaussian was centred at $(36.4 \pm 0.6 \text{ nm})^{-1}$, with amplitude 296 ± 31 and half-width $(4.25 \pm 0.53) \times 10^{-3} \text{ nm}^{-1}$. The centre corresponds to the position of the first actin layer line²⁹. There is a hint of a peak at the position of the second actin layer line at $\sim (18.5 \text{ nm})^{-1}$, also seen on the pattern in Fig. 3. The background was fitted by a linear function $-(192 \pm 17) + (1,600 \pm 186) \times S$, where S is the reciprocal spacing (in nm^{-1}). The decrease in background at the higher temperature is seen as negative intensity in the differential pattern. The increase in the I_{1a} intensity with temperature was seen in both outer $(12-8) \text{ nm}^{-1}$ and inner $(35-12) \text{ nm}^{-1}$ regions of radial integration. The parameters of gaussian fitting are presented as mean $\pm$ s.e.m.

tilting the structural transition responsible for force generation, or merely a consequence of the change in force induced by length perturbations²⁷? The absence of a decrease in $I_{14.5}$ after a 1.7-fold tension rise induced by the temperature jump suggests that it is the latter. $\square$

Methods

Muscle fibres. Muscle fibres from m. semitendinosus of *Rana temporaria* were permeabilized with Triton X-100 and slightly crosslinked with EDC (1-ethyl-(3-(3-dimethylamino)propyl)carbodiimide) to preserve structural order, as described²¹. From 10–20% of myosin heads were crosslinked with EDC²¹, allowing 90–80% of the heads to cycle and generate force in a normal way. The fibres were activated by a solution containing (in mM): MOPS 100; Ca-EGTA 10; ATP 5; Mg^{2+} 1; creatine phosphate 32; chicken creatine kinase²¹ 2.5–3 mg ml^{-1} , and suspended in air at 5–6 °C for 1 s before X-ray exposure.

X-ray diffraction. Experiments were done on beam lines 2.1 and 16.1 of the Synchrotron Radiation Source (CLRC Daresbury, UK) as described²¹. The intensities of the $I_{1,0}$, $I_{1,1}$ and $I_{14.5}$ reflections were collected from 478 repeats of the protocol in 13 fibres (10–50 repeats per fibre, until force decreased to 85%). Details of the Joule temperature jump (T-jump)⁸ (imposed by passing an a.c. voltage pulse through the fibre suspended in air) and X-ray camera²⁸ have been described. The X-ray data were analysed using BSL and XOTOKO software from the Daresbury Laboratory²¹. The off-meridional intensity of the first actin layer line (I_{1a}) was integrated at a radial reciprocal spacing of $(35 \text{ nm})^{-1}$ to $(8 \text{ nm})^{-1}$, where this layer line is seen above the background in rigor. We also studied temperature-induced changes in I_{1a} in the inner $((35 \text{ nm})^{-1}$ to $(12 \text{ nm})^{-1}$) and outer $((12 \text{ nm})^{-1}$ to $(8 \text{ nm})^{-1}$) radial regions to check whether the intensity changes of the layer line take place at different radii.

Received 10 March; accepted 29 April 1997.

- Goldman, Y. E. & Simmons, R. M. Active and rigor muscle stiffness. *J. Physiol., Lond.* **269**, 55P–57P (1977).
- Higuchi, H., Yanagida, T. & Goldman, Y. E. Compliance of thin filaments in skinned fibers of rabbit skeletal muscle. *Biophys. J.* **69**, 1000–1010 (1995).
- Haselgrove, J. C. & Huxley, H. E. X-ray evidence for radial cross-bridge movement and for the sliding filament model in actively contracting skeletal muscle. *J. Mol. Biol.* **77**, 549–568 (1973).
- Huxley, H. E., Faruqi, A. R., Kress, M., Bordas, J. & Koch, M. H. J. Time resolved X-ray diffraction studies of the myosin layer line reflections during muscle contraction. *J. Mol. Biol.* **158**, 637–684 (1982).
- Yagi, N. Intensification of the first actin layer-line during contraction of frog skeletal muscle. *Adv. Biophys.* **27**, 35–43 (1991).
- Amemiya, Y., Wakabayashi, K., Tanaka, H., Ueno, Y. & Miyahara, J. Laser-stimulated luminescence used to measure x-ray diffraction of a contracting striated muscle. *Science* **237**, 164–168 (1987).
- Bordas, J. *et al.* Two-dimensional X-ray diffraction of muscle: recent results. *Adv. Biophys.* **27**, 15–33 (1991).
- Bershtsky, S. Y. & Tsaturyan, A. K. Force generation and work production by covalently cross-linked actin-myosin cross-bridges in rabbit muscle fibers. *Biophys. J.* **69**, 1011–1021 (1995).
- Huxley, H. E. The mechanism of muscular contraction. *Science* **164**, 1356–1366 (1969).
- Huxley, A. F. & Simmons, R. M. Proposed mechanism of force generation in striated muscle. *Nature* **233**, 533–538 (1971).
- Huxley, H. E. & Kress, M. Cross-bridge behaviour during muscle contraction. *J. Musc. Res. Cell Motility* **6**, 153–161 (1985).
- Hirose, K., Lenart, T. D., Murray, J. M., Franzini-Armstrong, C. & Goldman, Y. E. Flash and smash: rapid freezing of muscle fibers activated by photolysis of caged ATP. *Biophys. J.* **65**, 397–408 (1993).
- Ling, N., Shrimpton, C., Sleep, J., Kendrick-Jones, J. & Irving, M. Fluorescent probes of orientation of myosin regulatory light chains in relaxed, rigor and contracting muscle. *Biophys. J.* **70**, 1836–1846 (1996).
- Ostap, E. M., Barnett, V. A. & Thomas, D. D. Resolution of three structural states of spin-labelled myosin in contracting muscle. *Biophys. J.* **69**, 177–188 (1995).
- Kress, M., Huxley, H. E., Faruqi, A. R. & Hendrix, J. Structural changes during activation of frog muscle studied by time-resolved X-ray diffraction. *J. Mol. Biol.* **188**, 325–342 (1986).
- Ford, L. E., Huxley, A. F. & Simmons, R. M. Tension responses to sudden length change in stimulated frog muscle fibres near slack length. *J. Physiol., Lond.* **269**, 441–515 (1977).
- Huxley, H. E. *et al.* Changes in the X-ray reflections from contracting muscle during rapid mechanical transients and their structural implications. *J. Mol. Biol.* **169**, 469–506 (1983).
- Irving, M., Lombardi, V., Piazzesi, G. & Ferenczi, M. A. Myosin head movements are synchronous with the elementary force-generating process in muscle. *Nature* **357**, 156–158 (1992).

19. Lombardi, V., Piazzesi, G., Ferenczi, M. A., Thirlwell, H. & Irving, M. Elastic distortion of myosin heads and repriming of the working stroke in muscle. *Nature* **374**, 553–555 (1995).
20. Piazzesi, G. *et al.* Changes in the x-ray diffraction pattern from single, intact muscle fibers produced by rapid shortening and stretch. *Biophys. J.* **68**, 925–985 (1995).
21. Bershtitsky, S. Y. *et al.* Mechanical and structural properties underlying contraction of skeletal muscle fibers after partial 1-ethyl-[3-(3-dimethylamino)propyl]carbodiimide cross-linking. *Biophys. J.* **71**, 1462–1474 (1996).
22. Vainstein, B. K. *Diffraction of X-rays by Chain Molecules* (Izdatelstvo Akademii Nauk SSSR, Moscow, 1963).
23. Cooke, R., Crowder, M. S., Wendt, C. H., Barnett, V. A. & Thomas, D. D. in *Contractile Mechanisms in Muscle* 413–423 (Plenum, New York, London, 1984).
24. Harford, J. J. & Squire, J. M. Evidence for structurally different attached states of myosin cross-bridges on actin during contraction of fish muscle. *Biophys. J.* **63**, 387–396 (1992).
25. Brenner, B. & Yu, L. C. Structural changes in the actomyosin cross-bridges associated with force generation. *Proc. Natl Acad. Sci. USA* **90**, 5252–5256 (1993).
26. Irving, M. *et al.* Tilting of the light-chain region of myosin during step length changes and active force generation in skeletal muscle. *Nature* **375**, 688–691 (1995).
27. Martin-Fernandez, M. A. *et al.* Time-resolved X-ray diffraction studies of myosin head movements in live frog sartorius muscle during isometric and isotonic contractions. *J. Musc. Res. Cell Motility* **15**, 319–348 (1994).
28. Towns-Andrews, E. *et al.* Time-resolved x-ray diffraction station: x-ray optics, detectors and data acquisition. *Rev. Scient. Instr.* **60**, 2346–2349 (1989).
29. Huxley, H. E. & Brown, W. The low-angle X-ray diagram of vertebrate striated muscle and its behaviour during contraction and rigor. *J. Mol. Biol.* **30**, 383–434 (1967).

Acknowledgements. We thank M. Irving for comments on the early version of the manuscript, D. R. Trentham for support, and E. Towns-Andrews and the non-crystalline-diffraction team at CLRC Daresbury Laboratory for hard- and software support. The work was supported by the MRC and grants from the EU, INTAS, ISF, HHMI, RFBR and by CCP13 of EPSRC and BBSRC.

Correspondence and requests for materials should be addressed to M.A.F. (e-mail: m-ferenc@nimr.mrc.ac.uk).

Inhibition of death receptor signals by cellular FLIP

Martin Irmier^{*,†}, Margot Thome^{*,†}, Michael Hahne^{*}, Pascal Schneider[‡], Kay Hofmann[‡], Véronique Steiner^{*}, Jean-Luc Bodmer^{*}, Michael Schröter^{*}, Kim Burns^{*}, Chantal Mattmann^{*}, Donata Rimoldi[§], Lars E. French^{||} & Jürg Tschopp^{*}

^{*} Institute of Biochemistry, [§] Ludwig Institute of Cancer Research, Lausanne branch, University of Lausanne, and [‡] the Swiss Institute for Experimental Cancer Research (ISREC), BIL Biomedical Research Center, Chemin des Boveresses 155, CH-1066 Epalinges, Switzerland
^{||} Department of Dermatology, University of Geneva, Medical School, CH-1211 Geneva 4, Switzerland

[†] These authors contributed equally to this work.

The widely expressed protein Fas is a member of the tumour necrosis factor receptor family which can trigger apoptosis¹. However, Fas surface expression does not necessarily render cells susceptible to Fas ligand-induced death signals^{1,2}, indicating that inhibitors of the apoptosis-signalling pathway must exist. Here we report the characterization of an inhibitor of apoptosis, designated FLIP (for FLICE-inhibitory protein), which is predominantly expressed in muscle and lymphoid tissues. The short form, FLIP_S, contains two death effector domains and is structurally related to the viral FLIP inhibitors of apoptosis³, whereas the long form, FLIP_L, contains in addition a caspase-like domain in which the active-centre cysteine residue is substituted by a tyrosine residue. FLIP_S and FLIP_L interact with the adaptor protein FADD^{4,5} and the protease FLICE^{6,7}, and potentially inhibit apoptosis induced by all known human death receptors¹. FLIP_L is expressed during the early stage of T-cell activation, but disappears when T cells become susceptible to Fas ligand-mediated apoptosis. High levels of FLIP_L protein are also detectable in melanoma cell lines and malignant melanoma tumours. Thus FLIP may be implicated in tissue homeostasis as an important regulator of apoptosis.

We have recently described a family of six viral inhibitors (v-FLIPs) that are present in several γ -herpesviruses and molluscipoxvirus, and which contain two death effector domains (DEDs)³.

These inhibitors block the early signalling events of the death receptors Fas, TRAMP (wsl/DR-3/Apo-3), TRAIL-R (DR-4) and tumour necrosis factor receptor 1 (TNF-R1) (refs 1, 3, 8). v-FLIPs bind to the DED of FADD (MORT-1) and interfere with the FADD–FLICE (Caspase 8/MACH/Mch-5) interaction, thereby inhibiting the recruitment and activation of FLICE by Fas³. Some v-FLIPs may also exert their inhibitory effect by binding directly to the DEDs of FLICE⁸. Because many viral genes are adopted host genes, we considered the possibility that cellular homologues of the v-FLIPs may exist. We screened public databases with a generalized profile⁹ constructed from the six known members of the v-FLIP family and identified a partial human sequence with highly significant homology ($P < 10^{-10}$) to v-FLIPs (Fig. 1a). When the corresponding expressed sequence tag (EST) clone was used to screen a cDNA library from activated human peripheral blood leukocytes, we isolated several cDNA clones encoding a protein with an overall structural organization similar to the v-FLIPs, which we designated FLIP_S. This protein consists of two DEDs, followed by a carboxy-terminal extension of approximately 50 amino acids. An alternatively spliced cDNA coding for a longer form of FLIP, designated FLIP_L, was also isolated. Within this protein the two DEDs are followed at the C terminus by a caspase-like domain, giving rise to a molecule structurally similar to FLICE (Fig. 1a). The gene for FLIP co-localizes with the FLICE homologue Mch4 (caspase 10)¹⁰ to the chromosome 2q33 (ref. 11), suggesting that the genes coding for Mch4, FLIP and probably FLICE arose by gene duplication. However, the amino-acid position corresponding to the active-site cysteine of FLICE is substituted by a Tyr residue in FLIP_L. The active-site cysteine was absent in all human FLIP_L-encoding cDNA clones analysed (four clones), as well as in mouse FLIP_L, which had been isolated by screening a murine heart muscle cDNA library (Fig. 1a). Moreover, of the two other residues involved in catalysis^{12,13} and conserved in members of the caspase family, only the Gly residue is conserved (Fig. 1a); the His residue is substituted by an Arg (human FLIP_L) or a Leu residue (mouse FLIP_L). FLIP_L was therefore expected to be proteolytically inactive, similar to the FLICE-1 (ref. 7) and FLICE-2 (ref. 14) mutants in which the catalytic cysteine residue had been replaced. Indeed, in contrast to FLICE, the caspase region of FLIP_L was unable to cleave two caspase-specific peptide substrates⁷ (Fig. 1b). However, we cannot completely exclude the possibility that the caspase-like region of FLIP_L can act on different peptides.

Northern blots of human tissues revealed that four main FLIP RNA species exist which are predominantly expressed in heart, skeletal muscle and peripheral blood leukocytes (PBLs) (Fig. 1c). The shortest transcript (1.1–1.3 kilobases) probably corresponds to FLIP_S, as it was not detected with an RNA probe uniquely spanning the caspase region of FLIP_L. Three longer, less abundantly expressed transcripts reacted with both the caspase and the DED probes, and presumably corresponded to FLIP_L.

Triggering of Fas results in the incorporation of v-FLIPs into the death-inducing signalling complex (DISC)^{3,15}. To assess whether cellular FLIP was functionally similar to v-FLIP, we analysed its interaction with FADD, which connects Fas with FLICE during death signal transduction^{6,7}. Indeed, both human FLIP_L (relative molecular mass (M_r) of ~55K) and FLIP_S (M_r ~28K) associated strongly with FADD (Fig. 2a) when coexpressed in 293T cells. The FLIP_L–FADD interaction was mediated by the DEDs of FLIP_L, as the caspase-like region of FLIP_L (FLIP_L) displayed no affinity for FADD. In cells transfected with expression vectors for FLIP_S, FLIP_L, FADD and the cytoplasmic portion of Fas, a stable Fas–FADD–FLIP_L (or FLIP_S) complex was formed, indicating that both forms of FLIP can incorporate into the DISC of Fas.

Current evidence suggests that precursor caspases are activated and autoprocessed as a dimeric complex, ultimately resulting in a stable (p10/p20)₂ complex^{12,13,16}. Correspondingly FLICE overexpression in 293T cells resulted in the formation of dimers or

higher oligomers of FLICE (the experimental conditions do not allow the distinction between these two possibilities) (Fig. 2b). We therefore anticipated that the FLICE-like FLIP_L would also form homodimers. However, FLIP_L dimers (or oligomers) were not detectable (Fig. 2c). In contrast, FLIP_L interacted with FLICE (Fig. 2d), indicating that heterodimers can be formed between functional (FLICE) and non-functional (FLIP_L) caspases. The presence of FLICE in immunoprecipitates of FLIP_S and of FLIP_C (the caspase-like region of FLIP_L) revealed that both the N-terminal DEDs, as well as the C-terminal caspase-like region of FLIP_L, contribute to the interaction with FLICE (Fig. 2d).

Overexpression of FLIP_L in 293T cells in the absence of z-VAD-fmk (a caspase inhibitor which had been included in the experiments described above) resulted in the appearance of two molecular species. One of these corresponds to the expected molecular size of full-length FLIP_L ($M_r \sim 55K$), and the second is a C-terminally truncated form of FLIP_L ($M_r \sim 43K$; FLIP₄₃), which arises from cleavage after Asp 376 of FLIP_L (data not shown) corresponding to the p17–p12 boundary in other related caspase members (Figs. 1a, 2e). Initial processing of FLICE during Fas-mediated activation

occurs at the corresponding site¹⁷. Precursor FLICE is a natural substrate of active FLICE^{6,17}, and it is therefore likely that the FLICE-like FLIP_L is processed by endogenous FLICES. Processing of full-length FLIP_L into the 43K form was also observed when Raji B cells, stably transfected with FLIP_L (see below), were treated with soluble Fas ligand (sFasL) (Fig. 2e). The cleavage of FLIP_L is reminiscent of the cleavage of the viral apoptosis inhibitor p35 (refs 18, 19), which is an excellent substrate for several caspases, yet its cleavage results in a tighter binding to the caspases. Similarly, FLIP_L cleavage through receptor-activated FLICE may result in a strong FLIP_L–FLICE interaction, thereby blocking further processing and activation of FLICE. This model, together with the FLIP interaction data (Fig. 2), predicts that FLIP is capable of arresting death receptor signals. This was validated experimentally in 293T cells, where the extensive cellular death induced by the overexpression of Fas was efficiently reduced by coexpression of FLIP_L (Fig. 3a) and FLIP_S (data not shown). Moreover, coexpression of FLIP_S or FLIP_L, but not of FLIP_C (the caspase-like domain of FLIP_L; see Fig. 1a), protected 293T cells from apoptosis induced by TRAMP (DR3/wsl/Apo-3)^{20,23} (Fig. 3b), indicating that the presence of the DEDs within FLIP_L is essential for

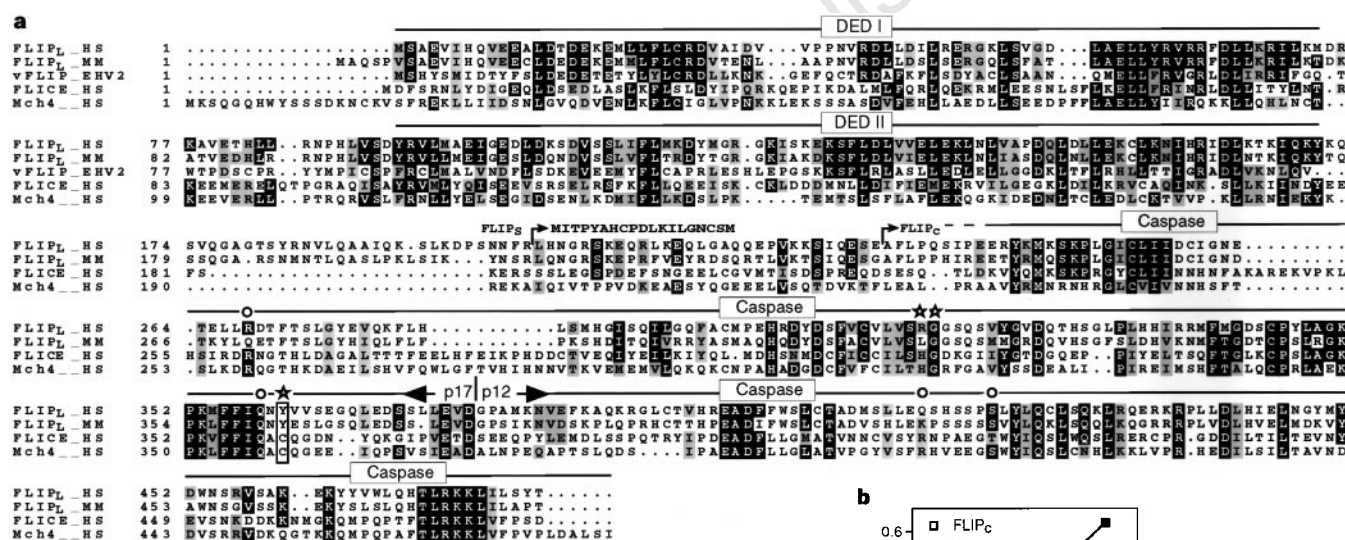
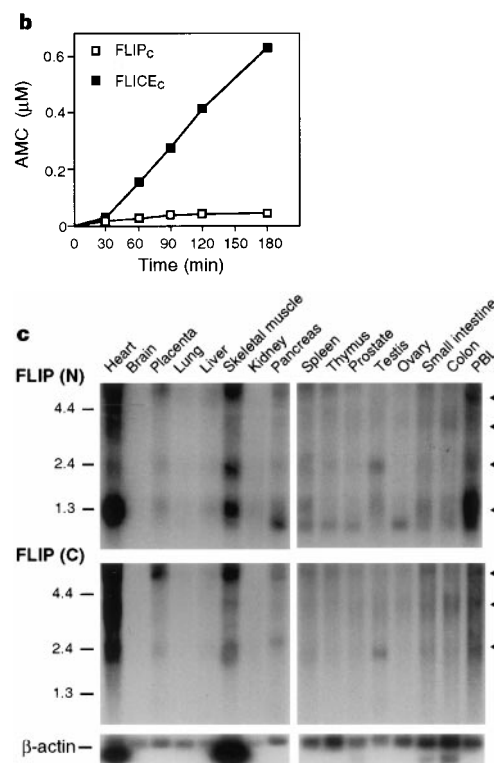


Figure 1 **a**, Amino-acid sequence alignment of human (HS) and mouse (MM) FLIPs, the viral FLIP (v-FLIP) from equine herpesvirus-2 (EHV2) (GenBank accession no. U20824), FLICE (ref. 6) and Mch4 (ref. 10). The long form of FLIP (FLIP_L) contains two DEDs and a caspase-like domain, linked by a region of ~65 amino acids. The sequence of the alternatively spliced short form of FLIP (FLIP_S) includes the two DEDs. Thus, the first 202 amino acids of the long and the short form of human FLIP are identical; only the unique C-terminal 19 amino acids of FLIP_S are shown here. A typical splice consensus sequence is found at the site of divergence between FLIP_L and FLIP_S. For each block of aligned sequences, black boxes indicate >50% amino-acid sequence identity; while grey shading indicates >50% sequence similarity through conservative amino-acid substitutions. The positions corresponding to the residues of FLICE and Mch4 involved in catalysis^{12,13} (His, Gly and Cys) are marked by a star, and the active-centre Cys is boxed. Open circles indicate caspase residues constituting the binding pocket for the carboxylate chain of the P1 Asp residue^{12,13}. The boundary of the p17 and p12 subunits of FLICE^{6,7} and the start of the caspase region-containing construct of FLIP_L (FLIP_C) are indicated. **b**, The caspase homology region of FLIP_L lacks protease activity. Cleavage kinetics of the poly(ADP ribose) polymerase (PARP)-derived fluorogenic substrate, Ac-DEVD-AMC (50 μM) by the caspase homology region of FLICE (Ser 217 through the C terminus; filled symbols) and of FLIP_L (Ala 233 through the C terminus; open symbols). Both proteins failed to cleave the fluorogenic substrate Ac-YVAD-AMC (ICE-cleavage site in the interleukin-1β precursor; data not shown). **c**, Tissue distribution of the FLIP transcripts. Northern blots (2 μg poly(A⁺)RNA per lane) of various human tissues were probed with antisense RNA covering the N-terminal sequence of FLIP_L containing the two DEDs (FLIP(N), top) and the C-terminal caspase-like domain (FLIP(C), bottom), respectively. Arrowheads indicate FLIP transcripts detected only by the FLIP(N) probe (open arrows) or by both probes (FLIP(N) and FLIP(C), filled arrows). The blots were subsequently hybridized with a β-actin probe.



the protective function of FLIP_L. FLIP_L and FLIP_S also protected 293T cells against apoptosis induced by overexpression of the cytoplasmic domain of TNF-R1 (ref. 24), albeit less efficiently (Fig. 3c and data not shown). Surprisingly, the transfection of high concentrations of FLIP_L expression vector alone (but not of FLIP_S) led to spontaneous cell death in 293T cells (Fig. 3d), raising the possibility that expression levels of FLIP_L may determine the life or death of a cell. We therefore stably transfected Raji B and Jurkat T cell lymphomas with FLIP_L or FLIP_S (Fig. 3e, f). In all clones obtained (and melanomas and T cells; see below), FLIP protein levels were at least 10-fold lower than the levels achieved in FLIP-expressing 293T cells undergoing apoptosis (Fig. 3e), suggesting that the extremely high concentration of FLIP observed in transfected 293T cells may not be physiological. The amount of FLIP protein present in stably transfected lymphocyte clones correlated well with the relative resistance to sFasL-induced cell death (Fig. 3e, f). Western blot analyses of FLIP_S and FLIP_L in cells that showed equal levels of protection against apoptosis consistently revealed that FLIP_L was a more potent inhibitor of cell death than was FLIP_S (Fig. 3e, f). In the Jurkat clone JFS5, for example, the relative level of FLIP_S protein was higher than the level of FLIP_L expressed in Jurkat clone JFL2 or JFL1. However, only the FLIP_L-expressing clones were protected against sFasL-induced apoptosis (the Fas surface expression was identical in all FLIP-expressing clones; data not shown). In contrast, the FLIP_S levels of JFS5 were sufficient to partly protect the clone against sTRAIL-induced apoptosis (Fig. 3e), indicating that, relative to Fas, the TRAIL-R signalling pathway is more effectively blocked. In view of the recent observation that TRAIL-R does not recruit FADD²⁵, this suggests that FLIP may preferentially interact with an unknown DED-containing protein specifically recruited by the TRAIL-R or, alternatively, that a second FADD-interacting

TRAIL-R exists. FLIP expression did not modulate apoptosis triggered by various other stimuli, such as the kinase inhibitor staurosporine (Fig. 3g), or growth factor withdrawal (data not shown).

FLIP was able to protect lymphoma cell lines against death receptor-induced apoptosis, so we investigated whether FLIP was implicated in the resistance of T lymphocytes against FasL-mediated apoptosis observed during early stages of activation². In accordance with published data², Fas-expressing human T cells were resistant to sFasL-induced apoptosis one day after mitogen stimulation despite there being abundant surface-exposed Fas levels, whereas they became sensitive after an additional 5 days in culture (Fig. 4a). This transient resistance correlated with FLIP expression levels (Fig. 4a). At day one, a 55K anti-FLIP-reactive band was detected at a position identical to that found in FLIP_L-transfected, z-VAD-fmk-treated 293T cells. In FasL-responding T cells activated for 6 days, the level of expression of FLIP had decreased considerably. An anti-FLIP reactive protein migrating at ~28K (corresponding to FLIP_S) was not detected, despite the presence of high FLIP_S RNA levels in PBLs (see Fig. 1c). These results suggest that FLIP_S expression may be controlled at the post-transcriptional level. Similar results were obtained with mitogen (concanavalin A)-activated mouse splenocytes (Fig. 4b). Both the 55K and the processed 43K forms of FLIP_L were present in these cells at day one of stimulation, whereas FLIP_L was no longer detectable in FasL-sensitive lymphocytes analysed 3 days after stimulation (Fig. 4b). Thus activation of T cells induces a transient resistance to Fas-induced apoptotic signals that correlates with increased FLIP expression. FasL-resistant activated T cells fail to recruit and activate FLICE²⁶ as would be predicted if FLIPs played a major role in this protection.

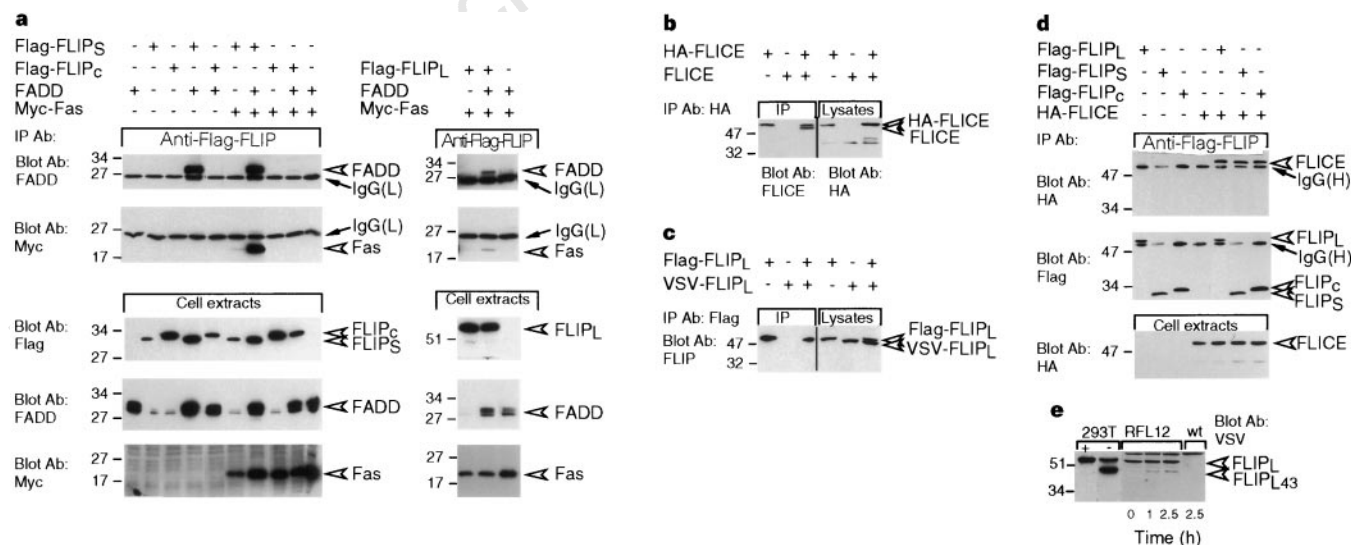
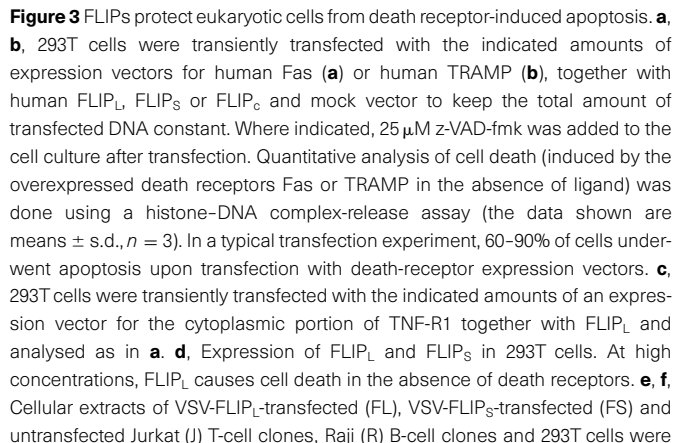


Figure 2 FLIPs interact with FADD and FLICE. **a**, 293T cells were transiently transfected with expression vectors encoding Flag-FLIP_S, Flag-FLIP_L, Flag-FLIP_c (covering the caspase-like region of FLIP_L), FADD and the myc-tagged cytoplasmic domain of Fas. In these and the following experiments, appropriate quantities of vector alone were added to keep the total amount of transfected plasmid constant and 25 μ M z-VAD-fmk was added to FLIP_L transfections (except in **e**). Cells were lysed 30 h after transfection, and anti-Flag immunoprecipitates (IP) or total cell extracts were analysed for the presence of FADD, Flag-FLIP or Myc-Fas by western blotting. **b**, Western blot analysis of anti-HA-FLICE immunoprecipitates and cell lysates from 293T cells, transfected with expression vectors encoding HA-FLICE and FLICE as indicated. **c**, Western blot analysis of anti-Flag immunoprecipitates and cell lysates from 293T cells transiently transfected with expression vectors encoding Flag-FLIP_L and VSV-FLIP_L. **d**, Anti-Flag-FLIP immunoprecipitates and cell extracts from Flag-FLIP_L, Flag-FLIP_S,

Flag-FLIP_c and HA-FLICE transfected 293T cells were analysed by anti-HA and anti-Flag western blotting. **e**, FLIP_L processing in sFasL-treated Raji cells. 293T cells were incubated in the absence (-) or presence (+) of the caspase inhibitor z-VAD-fmk. In the absence of the inhibitor, FLIP_L (55K) is partly processed to a 43K fragment (FLIP_L43) in 293T cells. A Raji B lymphoma clone stably transfected with N-terminally VSV-tagged FLIP_L (clone RFL12; see Fig. 3f) and the parental clone (Raji wt) were treated for the indicated times with 1 μ g ml⁻¹ of recombinant sFasL. Cells were lysed and processing of VSV-FLIP_L was analysed by anti-VSV western blotting 1 and 2.5 h after the addition of sFasL. The presence of sFasL leads to partial processing of FLIP_L into a 43K N-terminal fragment. A protein migrating at ~60K is detected nonspecifically. The FLIP_L-transfected Raji clone is resistant to apoptosis (see Fig. 3), whereas more than 90% of parental Raji cells have an apoptotic phenotype 2.5 h after the addition of sFasL.

Death receptor-induced apoptosis is an important event in tissue

homeostasis as exemplified by the accumulation of T cells in Fas-deficient mice¹. Therefore, a tight regulation of the balance between resistance and susceptibility of cells towards apoptosis is required. In T cells, this resistance is transient to avoid premature activation-induced cell death and to allow T cell-mediated help or cytotoxicity. In contrast, certain tumour cells seem to have lost the capacity to undergo apoptosis, either as a result of the loss of certain death receptors or by an impaired signalling pathway. Unlike the Bcl-2 family members which are potent inhibitors of apoptosis induced by growth factor withdrawal and γ -irradiation, FLIP seems to primarily block apoptosis induced by death receptors. In the examples analysed here (T cells and melanomas), FLIP expression correlates with the transient or acquired resistance to apoptosis, suggesting that FLIP may contribute to the blockade of the death signalling pathways in these cells. Highest FLIP mRNA levels are detected in the heart. The administration of anti-Fas antibodies into mice causes death from liver failure, not heart failure, despite the



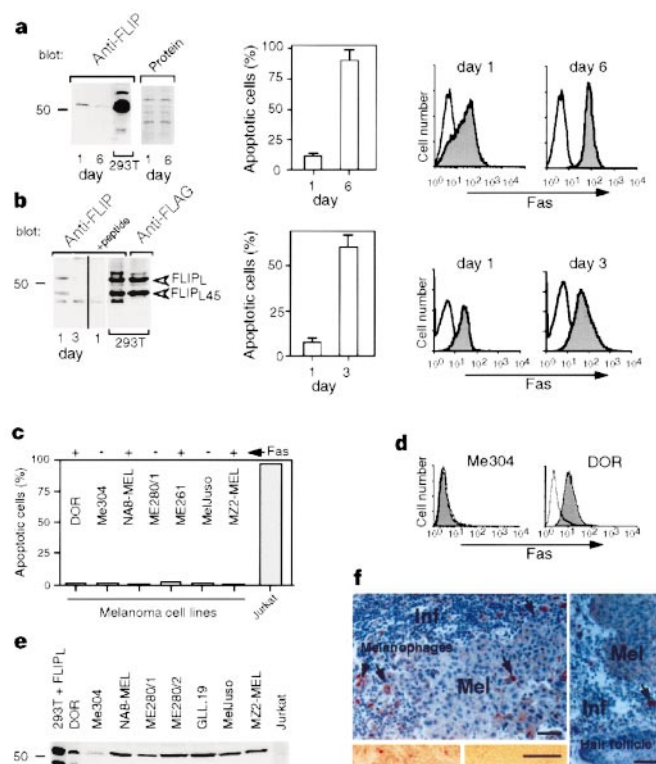


Figure 4 FLIP_L is expressed during T-cell activation and in human malignant melanoma. **a**, Western blot analysis of phytohaemagglutinin (PHA)-activated human T cells at day 1 and 6 of stimulation using the anti-FLIP AL109 antibody. A cell extract of FLIP_L-transfected (0.3 µg DNA) 293T cells (in the presence of z-VAD-fmk) is included as a control, and equal protein loading was verified by Ponceau S staining of the nitrocellulose blot. The middle panel shows the susceptibility of day 1 and 6 activated T cells to apoptosis induced by treatment with sFasL (0.2 µg ml⁻¹) for 6 h. The percentage of apoptotic cells was determined by the histone-DNA complex-release assay. All data are given as mean ± s.d. (n = 3). FACS analysis of Fas surface expression is shown on the right. Cells were stained with anti-Fas followed by an anti-mouse FITC-conjugated secondary antibody (filled curves) and analysed on a FACScan flow cytometer. Open curves represent cell staining with FITC-conjugated secondary antibody only. **b**, Western blot analysis of concanavalin A-activated mouse splenocytes at day 1 and 3 of stimulation. The quality of the anti-human FLIP antibody towards mouse FLIP was tested by comparing anti-FLIP and anti-Flag western blots of a cell extract from 293T cells transfected with a mouse Flag-FLIP_L expression vector in the absence of z-VAD-fmk. Moreover, in the western blot analysis of T cells from day 1, 10 µg ml⁻¹ of the peptide used for immunization was included as a specificity control. A band migrating at ~41K is detected nonspecifically and demonstrates equal protein loading. The middle panel shows the susceptibility of day 1 and 3 activated T cells to FasL-mediated apoptosis. Fas surface expression (right) was determined as in **a**, except that the biotinylated anti-mouse Fas antibody Jo2 in combination with streptavidin-FITC was used for the detection of the antigen. **c**, Melanoma cell lines resist killing by sFasL. Various cell lines were incubated with recombinant sFasL (1 µg ml⁻¹) for 6 h. The presence(+) or absence(-) of surface Fas was verified by FACS analysis (see Fig. 4d). FasL-sensitive Jurkat cells were used as control target cells. **d**, Fas surface expression on two selected melanoma cell lines (Me304 and DOR). Cells were stained as in **a**. **e**, Expression of FLIP_L in melanoma cell lines. Anti-FLIP western blot analysis of melanoma cell lines derived from several patients with malignant melanoma. Controls included 293T cells transfected with a FLIP_L expression vector and Jurkat cells. **f**, Expression of FLIP (brown) in two examples of human malignant melanoma (Mel, top and right panels). The tumours are surrounded by an inflammatory infiltrate (Inf). The dark cells (arrows) correspond to melanophages (macrophages with phagocytosed melanoma-derived melanin). The bottom sections from a metastatic melanoma were not counterstained with haematoxylin. The specificity of the reaction was assayed by the preincubation of the antibody with the peptide (10 µg ml⁻¹) used for immunization (bottom right). Scale bars, 20 µm each.

presence of high levels of Fas in the latter organ²⁸. Targeted disruption of the FLIP gene may therefore provide an insight into the physiological role of FLIP. □

Methods

Antibodies. Monoclonal antibodies used in immunoprecipitations and western blotting include: an anti-Flag antibody (Kodak International Biotechnologies); an anti-FADD antibody directed against amino acids 94–208 of human FADD (Transduction Laboratories); and antibodies against the myc epitope (9E10, Sigma), the VSV epitope (P5D4, Sigma) and the HA epitope (12CA5, from R. Iggo, Lausanne). An antiserum against FLICE (AL105) was generated using a peptide spanning amino acids 380–414 of FLICE (QTRYIPDEADFLGMATVNNVCYSYRN), and an antiserum against human FLIP (AL109) was generated against a peptide spanning amino acids 2–26 (SAEVIHQVEEALDTDEKEMFLCRD), synthesized using the multiple antigen technology²⁹. The antisera were affinity purified on the corresponding peptides coupled to CNBr-Sepharose 4B (Pharmacia). Antibodies used to detect Fas surface expression by FACScan analysis included the anti-human Fas antibody Apo-1-3 (Alexis, Dan Diego) and the anti-mouse Fas antibody Jo2 (Pharmingen).

Expression vectors, cell lines, melanoma and reagents. Expression vectors for FLIP_S, FLIP_L and FLIP_C (see Fig. 1a) were generated by PCR amplification and subcloning into PCR-3-derived vectors conferring an N-terminal Flag- or VSV-tag to the expressed proteins. An expression vector for the VSV-tagged cytoplasmic domain of the murine tumour necrosis factor receptor (TNF-R1) was generated by insertion of a PCR fragment corresponding to amino acids

242–454 in frame with an N-terminal VSV epitope into a vector derived from PCR-3 (Invitrogen). The expression vector for HA-tagged FLICE was a gift from J. P. Medema and M. E. Peter (Heidelberg, Germany). Recombinant soluble FasL was from Alexis (San Diego). Additional expression vectors, cell lines and reagents used in this study have been described previously^{3,20}. Melanoma cell lines were derived from tumour samples obtained from melanoma patients by surgical excision²⁷. Metastatic melanomas of patients were removed by surgery and snap-frozen in liquid nitrogen-cooled isopentane. Sections (10 µm thick) were cut and analysed as previously described²⁷. **Screening of cDNA libraries and northern blot analysis.** cDNA clones for human FLIP_S and FLIP_L were isolated from a cDNA library of activated PBLs (a gift of H. Eibel, Freiburg) using a ³²P-labelled DNA fragment (random primed labelling kit, Boehringer) encoding amino acids 1–170 of human FLIP_S and generated from EST clone 309776. A cDNA clone for murine FLIP_L was isolated from a murine heart muscle cDNA library (Stratagene) using the same probe. Northern blot analysis was performed by using human multiple tissue northern blots I and II (Clontech #7760-1 and #7759-1) according to the manufacturer's instructions. Antisense RNA probes containing the nucleotides corresponding to amino acids 1–170 (probe FLIP-N), 229–480 (FLIP-C) of FLIP_L, and β-actin were used.

Cell transfection, co-immunoprecipitation and western blotting. Stable transfections of Jurkat and Raji cells, transient transfection of 293T cells, immunoprecipitations and western blot analysis were done as previously described³. Jurkat and Raji clones or 293T cells were checked for expression of transfected proteins by anti-tag western blot analysis of postnuclear cell lysates with equivalent protein content.

Preparation and stimulation of T cells. Isolation of splenocytes and enriched human T cells and their stimulation with concanavalin A and PHA were done as described^{2,30}.

Apoptosis and cell proliferation assays. Reduced cell proliferation or apoptosis as a result of death receptor stimulation or overexpression was measured using the Celltiter 96 AQ proliferation assay (Promega) and the cell death detection ELISA (Boehringer Mannheim), respectively, as previously described³.

Protease activity assays. Flag fusion proteins of the caspase homology region of FLIP_L and FLICE were generated in bacteria using modified pQE16 (Qiagen) bacterial expression vectors. Protease activity was determined on fluorogenic substrates as described⁷.

Chromosomal localization. A sequence-tagged site¹¹ (A008B37) corresponding to FLIP maps to human chromosome 2 in the interval between the markers D2S116–D2S307. On the cytogenetic map, this interval corresponds to 2q33.

Received 19 May; accepted 11 June 1997.

- Nagata, S. Apoptosis by death factor. *Cell* **88**, 355–365 (1997).
- Klas, C., Debatin, K. M., Jonker, R. R. & Krammer, P. H. Activation interferes with the APO-1 pathway in mature human T cells. *Int. Immunol.* **5**, 625–630 (1993).
- Thome, M. *et al.* Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. *Nature* **386**, 517–521 (1997).
- Boldin, M. P. *et al.* A novel protein that interacts with the death domain of Fas/APO1 contains a sequence motif related to the death domain. *J. Biol. Chem.* **270**, 7795–7798 (1995).
- Chinnaiyan, A. M., O'Rourke, K., Tewari, M. & Dixit, V. M. FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. *Cell* **81**, 505–512 (1995).
- Muzio, M. *et al.* FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
- Boldin, M. P., Goncharov, T. M., Goltsev, Y. V. & Wallach, D. Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* **85**, 803–815 (1996).
- Bertin, J. *et al.* Death effector domain-containing herpesvirus and poxvirus proteins inhibit both Fas- and TNFR1-induced apoptosis. *Proc. Natl Acad. Sci. USA* **94**, 1172–1176 (1997).
- Bucher, P., Karplus, K., Moeri, N. & Hofmann, K. A flexible search technique based on generalized profiles. *Comput. Chem.* **20**, 3–24 (1996).
- Fernandes-Alnemri, T. *et al.* In vitro activation of CPP32 and Mch3 by Mch4, a novel human apoptotic cysteine protease containing two FADD-like domains. *Proc. Natl Acad. Sci. USA* **93**, 7464–7469 (1996).
- Schuler, G. D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
- Walker, N. P. C. *et al.* Crystal structure of the cysteine protease interleukin-1 β -converting enzyme—a (p20/p10)(2) homodimer. *Cell* **78**, 343–352 (1994).
- Wilson, K. P. *et al.* Structure and mechanism of interleukin-1 β converting enzyme. *Nature* **370**, 270–275 (1994).
- Vincenz, C. & Dixit, V. M. Fas-associated death domain protein interleukin-1 β -converting enzyme 2 (FLICE2), an ICE/Ced-3 homologue, is proximally involved in CD95- and p55-mediated death signaling. *J. Biol. Chem.* **272**, 6578–6583 (1997).
- Kischkel, F. C. *et al.* Cytotoxicity-dependent Apo-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.* **14**, 5579–5588 (1995).
- Gu, Y. *et al.* Interleukin-1 β converting enzyme requires oligomerization for activity of processed forms in vivo. *EMBO J.* **14**, 1923–1931 (1995).
- Medema, J. P. *et al.* FLICE is activated by association with the CD95 death-inducing signaling complex (DISC). *EMBO J.* **16**, 2794–2804 (1997).
- Bump, N. J. *et al.* Inhibition of ICE family proteases by baculovirus antiapoptotic protein p35. *Science* **269**, 1885–1888 (1995).
- Xie, D. & Horvitz, H. R. Inhibition of the *Caenorhabditis elegans* cell-death protease CED-3 by a CED-3 cleavage site in baculovirus p35 protein. *Nature* **377**, 248–251 (1995).
- Bodmer, J. L. *et al.* TRAMP, a novel apoptosis-mediating receptor with sequence homology to tumor necrosis factor receptor 1 and Fas/Apo-1/CD95. *Immunity* **6**, 79–88 (1997).
- Marsters, S. A. *et al.* Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF- κ B. *Curr. Biol.* **6**, 1669–1676 (1996).
- Kitson, J. *et al.* A death-domain-containing receptor that mediates apoptosis. *Nature* **384**, 372–375 (1996).
- Yu, G. L. *et al.* Signal transduction by DR3, a death domain-containing receptor related to TNFR-1 and CD95. *Science* **274**, 990–992 (1996).
- Tartaglia, L. A., Rothe, M., Hu, Y. F. & Goeddel, D. V. Tumor necrosis factor's cytotoxic activity is signaled by the p55 TNF receptor. *Cell* **73**, 213–216 (1993).
- Pan, G. *et al.* The receptor for the cytotoxic ligand TRAIL. *Science* **276**, 111–113 (1997).
- Peter, M. E. *et al.* Resistance of cultured peripheral T cells towards activation-induced cell death involves a lack of recruitment of FLICE (MACH/caspase 8) to the CD95 death-inducing signaling complex. *Eur. J. Immunol.* **27**, 1207–1212 (1997).
- Hahne, M. *et al.* Melanoma cells express Fas (Apo-1/CD95) ligand: Implications for tumor immune escape. *Science* **267**, 1363–1366 (1996).
- Nagata, S. & Golstein, P. The Fas death factor. *Science* **267**, 1449–1456 (1995).
- Francis, M. J. *et al.* Immunological evaluation of the multiple antigen peptide (MAP) system using the major immunogenic site of foot-and-mouth disease virus. *Immunology* **73**, 249–254 (1991).
- Lowin, B., Hahne, M., Mattmann, C. & Tschopp, J. Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways. *Nature* **370**, 650–652 (1994).

Acknowledgements. We thank S. Belli for reading this manuscript; C. Servi for peptide synthesis; and R. Bullani, S. Hertig, M. Rousseaux and T. Bormand for technical assistance. This work was supported by grants of the Swiss National Science Foundation (J.T.) and the European Molecular Biology Organisation (to M.T.).

Correspondence and requests for material should be addressed to J.T. (e-mail: jurg.tschopp@ib.unil.ch). The Genbank accession numbers for FLIP cDNA sequences are U97074 (human FLIP_L), U97075 (human FLIP_S), and U97076 (mouse FLIP_L).

Tom5 functionally links mitochondrial preprotein receptors to the general import pore

Klaus Dietmeier*, Angelika Hönlinger*, Ulf Bömer, Peter J. T. Dekker, Christoph Eckerskorn†, Fritz Lottspeich†, Michael Kübrich & Nikolaus Pfanner

Institut für Biochemie und Molekularbiologie, Universität Freiburg,

Hermann-Herder-Strasse 7, D-79104 Freiburg, Germany

† Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany

* These authors contributed equally to this study.

Most mitochondrial proteins are synthesized as preproteins on cytosolic polysomes and are subsequently imported into the organelle^{1–3}. The mitochondrial outer membrane contains a multisubunit preprotein translocase (Tom) which has receptors on the cytosolic side and a general import pore (GIP) in the membrane. Tom20–Tom22 and Tom70–Tom37 function as import receptors^{4–7} with a preference for preproteins that have amino-terminal presequences or internal targeting information, respectively. Tom40 is an essential constituent of the GIP^{8,9}, whereas Tom6 and Tom7 modulate the assembly and dissociation of the Tom machinery^{10,11}. Here we report the identification of Tom5, a small subunit that has a crucial role importing preproteins destined for all four mitochondrial subcompartments. Tom5 has a single membrane anchor and a cytosolic segment with a negative net charge, and accepts preproteins from the receptors and mediates their insertion into the GIP. We conclude that Tom5 represents a functional link between surface receptors and GIP, and is part of an 'acid chain'⁵ that guides the stepwise transport of positively charged mitochondrial targeting sequences.

For the preparation of Tom machinery, digitonin-lysed mitochondria from the yeast *Saccharomyces cerevisiae* were subjected to a coprecipitation with antibodies directed against Tom40. A non-characterized band (labelled 'Tom5' in Fig. 1a, lane 2) was subjected to N-terminal sequencing (Fig. 1b, sequence II). The sequence matched to a previously unknown open reading frame (ORF) on chromosome XVI (cosmid 9659, nucleotides 6,512–6,661), encoding a 50-residue protein of relative molecular mass (M_r) 5.98K (Fig. 1b) which we term Tom5. Its sequence does not reveal significant homology to any known protein. The C-terminal half of Tom5 contains a hydrophobic segment (residues 27–45) of sufficient length to function as a membrane anchor. This segment is flanked by positively charged residues that may interact with negatively charged headgroups of membrane phospholipids. The N-terminal portion of Tom5 carries a negative net charge (Fig. 1b).

Pure outer-membrane vesicles^{12,13} contained a prominent band with an electrophoretic mobility like that of Tom5 (Fig. 1a, lane 1), and N-terminal amino-acid sequencing confirmed that it was indeed Tom5 (Fig. 1b, sequence I). An antiserum directed against an N-terminal peptide of Tom5 was generated (Fig. 1a, lane 3). To assess whether Tom5 is an integral membrane protein, mitochondria were treated with sodium carbonate at pH 11.5 (ref. 11). This treatment allows the extraction of soluble and peripheral membrane proteins (Fig. 1c, columns 3 and 5) but leaves integral membrane proteins in the membrane sheets (Fig. 1c, columns 2 and 4). Tom5 was fully resistant to extraction by sodium carbonate (Fig. 1c, column 1), indicating that it is an integral membrane protein. Tom5 was also resistant to treatment of mitochondria with trypsin at all concentrations tested (Fig. 1d, lanes 2–6), even after lysis of the mitochondria with Triton X-100 (Fig. 1d, lane 7). Treatment of

mitochondria with proteinase K, however, led to complete loss of material reacting with the anti-Tom5 antibodies at low concentrations of the protease (Fig. 1d, lanes 9–13), demonstrating that the N terminus was digested. The marker proteins cytochrome *b*₂ of the intermembrane space and the ADP/ATP carrier of the inner membrane were not degraded (Fig. 1d, lanes 9–13), unless the mitochondrial membranes had been disrupted by detergent (Fig. 1d, lane 14), demonstrating that the outer-membrane barrier was not degraded by the protease treatment. We conclude that Tom5 is an integral mitochondrial outer-membrane protein that exposes its N terminus to the cytosolic side.

The Tom machinery can be separated into two subcomplexes, a 400K complex containing Tom40 and Tom22, and a 120K complex containing Tom70 (ref. 14). To determine which subcomplex contained Tom5, mitochondria were lysed with digitonin and protein complexes were analysed under non-denaturing conditions by blue native electrophoresis, followed by a second separation using urea-SDS-polyacrylamide gel electrophoresis. Tom5 migrated exclusively with the 400K complex (Fig. 1e, top). Together with the efficient coprecipitation of Tom5 with antibodies directed against Tom40 (Fig. 1a, lane 2), these results indicate that Tom5 is a component of the larger Tom subcomplex.

We constructed a yeast mutant with a deletion of the coding region of the *TOM5* gene. Growth of the *tom5Δ* strain at 30 °C was reduced in comparison to the wild type on both fermentable and non-fermentable carbon sources (threefold reduction of growth rate on glycerol-containing medium). At 37 °C, *TOM5* is essential for viability (Fig. 2a, sectors 2 and 5). Expression of *TOM5* from the centromeric vector pRS415 fully restored growth of *tom5Δ* cells (Fig. 2a, sectors 3 and 6), demonstrating that the growth defects were caused solely by the deletion of *TOM5*. Mitochondria were isolated from wild-type and *tom5Δ* cells and compared for their protein composition and ability to generate a membrane potential ($\Delta\psi$). Except for the lack of Tom5, the protein composition of *tom5Δ* mitochondria (analysed with marker proteins for all four subcompartments) was indistinguishable from that of wild-type

mitochondria (Fig. 2b). When digitonin-lysed *tom5Δ* mitochondria were separated by blue native electrophoresis, the stability of the 400K and 120K subcomplexes was not altered, as indicated by the unchanged mobility of Tom70, Tom40 and Tom22 (Fig. 1e, bottom). In contrast to Tom6 and Tom7 (refs 10, 11), Tom5 does not seem to modulate the stability of association of the Tom machinery. $\Delta\psi$ was assessed by use of the potential-sensitive fluorescent dye 3,3'-dipropylthiadicarbocyanine iodide (DiSC₃(5)). A decrease in fluorescence indicates the formation of a membrane potential^{15,16}. The degree of quenching was similar for *tom5Δ* and wild-type mitochondria (Fig. 2c), showing that *tom5Δ* mitochondria, like wild-type mitochondria, were able to generate a membrane potential.

³⁵S-labelled mitochondrial preproteins were imported into isolated *tom5Δ* and wild-type mitochondria. The following preproteins were tested: porin (from the outer mitochondrial membrane), cytochrome *c*₁ (intermembrane-space side of inner membrane), ADP/ATP carrier, F₁-ATPase subunit- β (matrix side of inner membrane), a fusion protein between the presequence of F₀-ATPase subunit 9 and dihydrofolate reductase (Su9-DHFR, targeted to the matrix)¹⁷, and the α -subunit of the mitochondrial-processing peptidase (matrix). Porin and the ADP/ATP carrier contain targeting information within the mature protein, whereas the other preproteins carry cleavable N-terminal presequences. Import of all preproteins was strongly reduced with *tom5Δ* mitochondria (Fig. 2d). We conclude that Tom5 is important for protein import in a manner independent of the type of targeting information and submitochondrial destination of preproteins.

We investigated whether Tom5 interacted with a preprotein in transit across the outer mitochondrial membrane. After accumulation of Su9-DHFR at the outer membrane of isolated mitochondria in the absence of a membrane potential^{15,2}, the amino-reactive homo-bifunctional crosslinking reagent ethylene glycolbis(succinimidyl-succinate) (EGS) was added. Several crosslinking products were generated (Fig. 3a, lanes 3–5 and 7–9). The cleavable preprotein was crosslinked to Tom20/Tom22 and to Tom40 (probably a

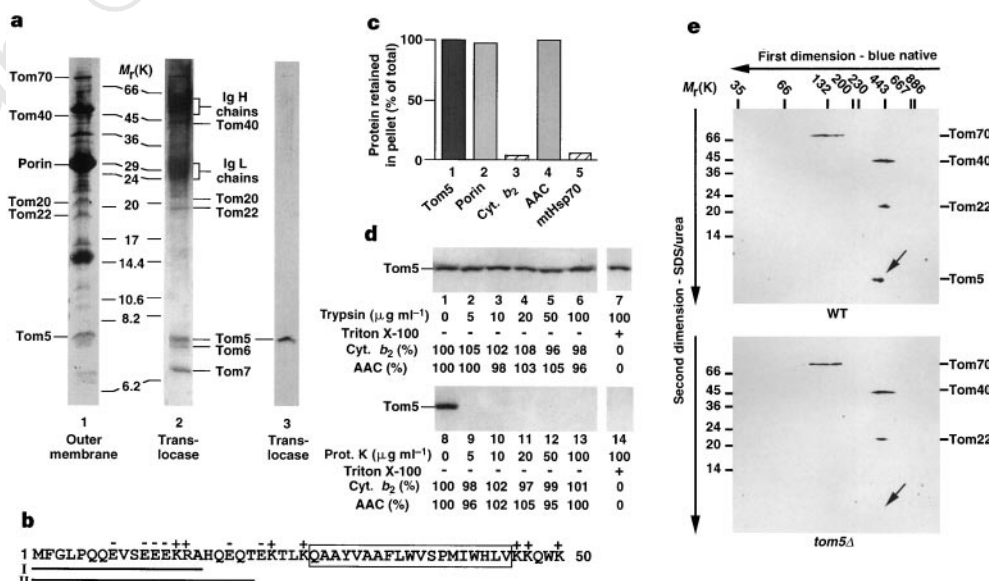


Figure 1 Tom5 is an integral mitochondrial outer membrane protein exposed to the cytosol. **a**, Purified outer-membrane vesicles (lane 1) and outer-membrane translocase (lane 2) stained with Coomassie brilliant blue R250. Immunoreaction of outer-membrane translocase with anti-Tom5N antibodies (lane 3). **b**, Primary structure of Tom5. N-terminal sequence of Tom5 derived from outer-membrane vesicles (I) or translocase (II) are underlined. The predicted membrane anchor is boxed. **c**, Tom5 is not extracted at pH 11.5. Cyt. *b*₂, cytochrome *b*₂; AAC, ADP/ATP carrier. **d**, The N terminus of Tom5 is located at the cytosolic side. Mitochondria were treated with trypsin and proteinase K (Prot. K) in the presence or absence of Triton X-100. **e**, Tom5 is present in the 400K Tom complex. Digitonin-lysed mitochondria were separated by blue native electrophoresis. WT, wild-type. Arrow, position of Tom5.

boxed. **c**, Tom5 is not extracted at pH 11.5. Cyt. *b*₂, cytochrome *b*₂; AAC, ADP/ATP carrier. **d**, The N terminus of Tom5 is located at the cytosolic side. Mitochondria were treated with trypsin and proteinase K (Prot. K) in the presence or absence of Triton X-100. **e**, Tom5 is present in the 400K Tom complex. Digitonin-lysed mitochondria were separated by blue native electrophoresis. WT, wild-type. Arrow, position of Tom5.

dimer of Tom40) as described^{5,12}. The major crosslinking product of ~35K was selectively recognized by antibodies directed against Tom5 (Fig. 3a, lane 2) and was absent in *tom5Δ* mitochondria (Fig. 3a, lanes 7–9) (Su9-DHFR is a 29K protein). We conclude that Tom5 is in close proximity to the accumulated preprotein. To

determine the step at which Tom5 functions in protein import, Su9-DHFR was arrested at different import stages^{5,12}. In the absence of a membrane potential, the precursor form of Su9-DHFR accumulated at the mitochondrial outer membrane (Fig. 3b, lanes 1 and 2). In the presence of a membrane potential, but at low levels

Figure 2 Tom5 is required for import of preproteins to all four mitochondrial subcompartments. **a**, Growth of yeast cells on fermentable (YPD) or non-fermentable (YPG) medium at 37°C. **b**, Mitochondrial content of marker proteins is comparable in wild-type (WT) and *tom5Δ*. Hsp60 (matrix); F₁-ATPase subunit-β (F₁β; matrix side of inner membrane). **c**, *tom5Δ* mitochondria are able to generate a membrane potential (Δψ) assessed by fluorescence spectrometry with the dye DiSC₃(5). **d**, Deletion of *TOM5* leads to severe import defects. ³⁵S-labelled preproteins were incubated with mitochondria at 25°C. The proteinase K-treated mitochondria were analysed by SDS-PAGE and digital autoradiography. The amount of protein imported into wild-type mitochondria after 5 min was set to 100% (bars, s.e.m. from 3–6 experiments). Precursor (p), intermediate (i) and mature (m) forms of the preproteins are shown; α-MPP, α-subunit of mitochondrial processing peptidase.

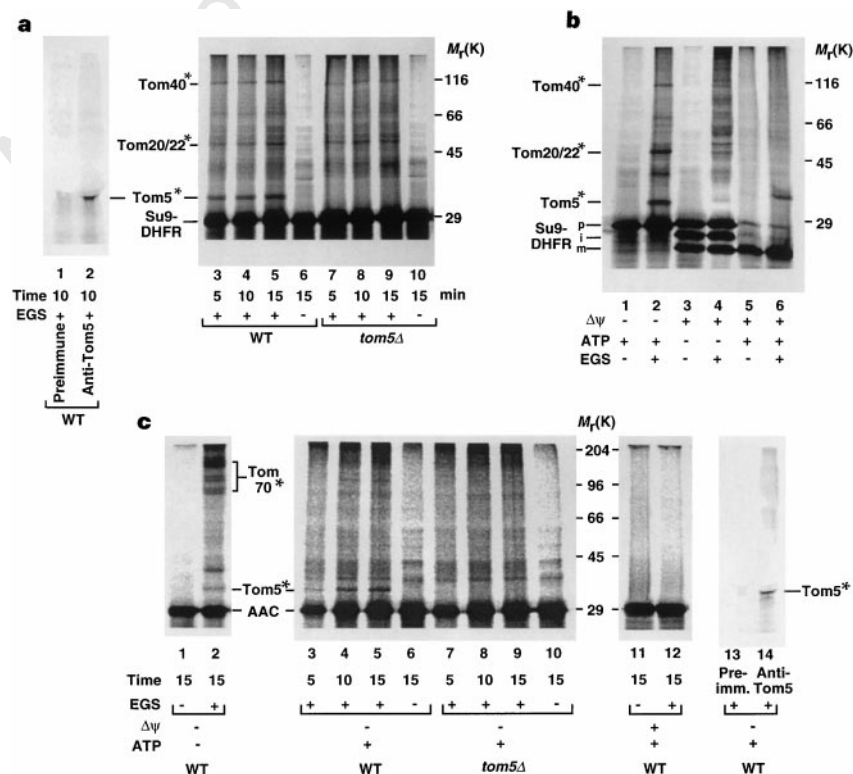
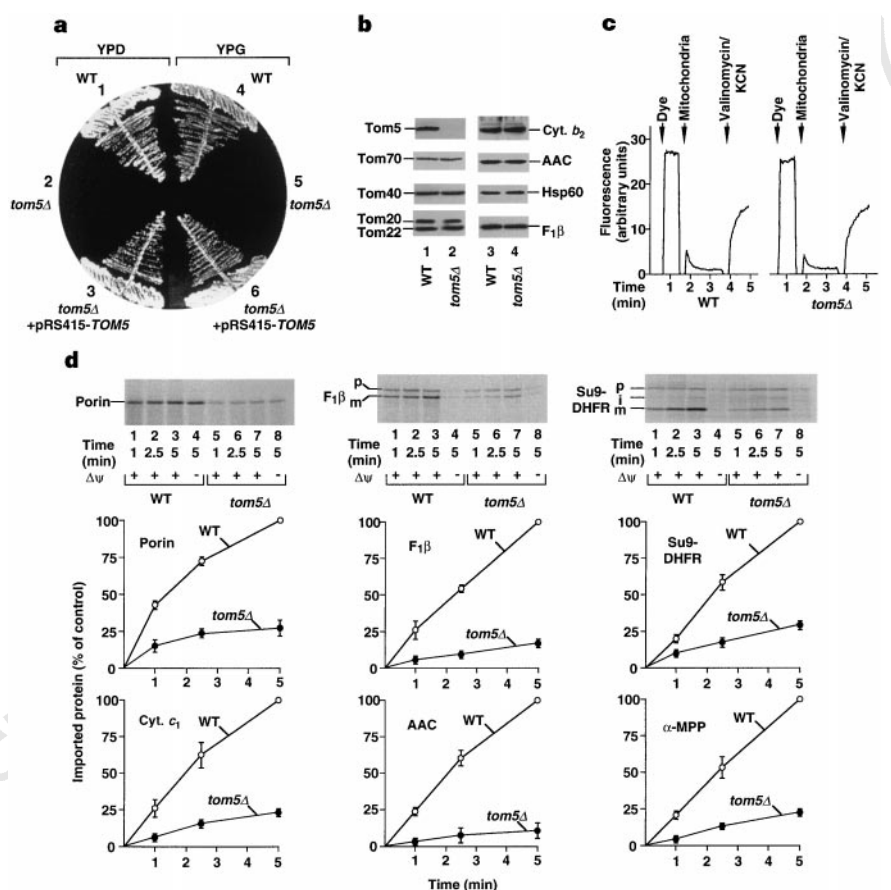


Figure 3 Crosslinking of preproteins to Tom5 depends on the import stage. **a**, ³⁵S-labelled precursor of Su9-DHFR was incubated with isolated de-energized mitochondria. Mitochondria were reisolated, subjected to crosslinking with EGS, and analysed directly (lanes 3–10) or subjected to immunoprecipitation (lanes 1 and 2). **b**, ³⁵S-labelled Su9-DHFR was incubated with wild-type (WT) mitochondria at different import stages¹² (outer membrane, lanes 1 and 2; membrane spanning, lanes 3 and 4; full import, lanes 5 and 6). **c**, ³⁵S-labelled precursor of the ADP/ATP carrier was accumulated at mitochondria at different import stages¹² (receptors, lanes 1 and 2; postreceptor, lanes 3–10, 13 and 14 (no protease treatment); full import, lanes 11 and 12). The efficiency of crosslinking to Tom5 (crosslinking product per accumulated preprotein) at the postreceptor stage was ~2.5-fold higher than that at the receptor stage. Asterisks indicate crosslinking products. Pre-imm., pre-immune.

of ATP, Su9-DHFR accumulated, spanning both outer and inner membranes, leading to partial cleavage of the presequence by the matrix-processing peptidase (Fig. 3b, lanes 3 and 4). In the presence of both a membrane potential and ATP, Su9-DHFR is imported into the mitochondrial matrix and processed to the mature form (Fig. 3b, lanes 5 and 6). Crosslinking between Su9-DHFR and Tom5 was efficient when the preprotein was accumulated at the outer membrane (Fig. 3b, lane 2), whereas only a weak crosslinking efficiency was observed with the two-membrane-spanning intermediate (Fig. 3b, lane 4). As expected, no crosslinking to Tom5 occurred when Su9-DHFR was imported into the matrix (Fig. 3b, lane 6).

We investigated whether Tom5 was also in close proximity to a non-cleavable preprotein with internal targeting information. In the absence of a membrane potential, the ADP/ATP carrier can be accumulated at two distinct stages at the outer membrane: at low levels of ATP, preferentially at receptor sites; and in the presence of ATP, preferentially at a post-receptor stage^{10,12,17}. In the presence of ATP, a crosslinking product between the preprotein and Tom5 was generated in wild-type mitochondria (Fig. 3c, lanes 3–5) that was recognized by antibodies directed against Tom5 (Fig. 3c, lane 14) and was absent in *tom5Δ* mitochondria (Fig. 3c, lanes 7–9). At low levels of ATP, prominent crosslinking occurred to Tom70 (mono-

meric and dimeric forms¹⁸), whereas crosslinking to Tom5 as weaker (Fig. 3c, lane 2). As expected, the ADP/ATP carrier imported to its functional location in the inner membrane was not crosslinked to Tom5 (Fig. 3c, lane 12). We conclude from this that Tom5 is in close proximity to both cleavable and non-cleavable preproteins that accumulated at the outer membrane. Crosslinking is most efficient at a postreceptor stage.

To determine which import stage requires Tom5, we prepared IgG directed against Tom5, and pre-bound them to isolated wild-type mitochondria. Subsequently, accumulation of the preprotein of the ADP/ATP carrier at distinct import stages was studied. Anti-Tom5 did not influence the binding of ADP/ATP carrier to receptor sites (Fig. 4a, column 3), but strongly inhibited the accumulation of ADP/ATP carrier at a postreceptor (GIP) stage of the outer membrane (Fig. 4a, column 6). Moreover, the preprotein Su9-DHFR was accumulated in the precursor form at the surface of *tom5Δ* mitochondria. When the mitochondria were pretreated with a low concentration of trypsin that removes the cytosolic receptor domains¹², this accumulation of pSu9-DHFR was strongly inhibited (Fig. 4b, lanes 2 and 4). We conclude that Tom5 is required for the import of cleavable and non-cleavable preproteins at a postreceptor stage.

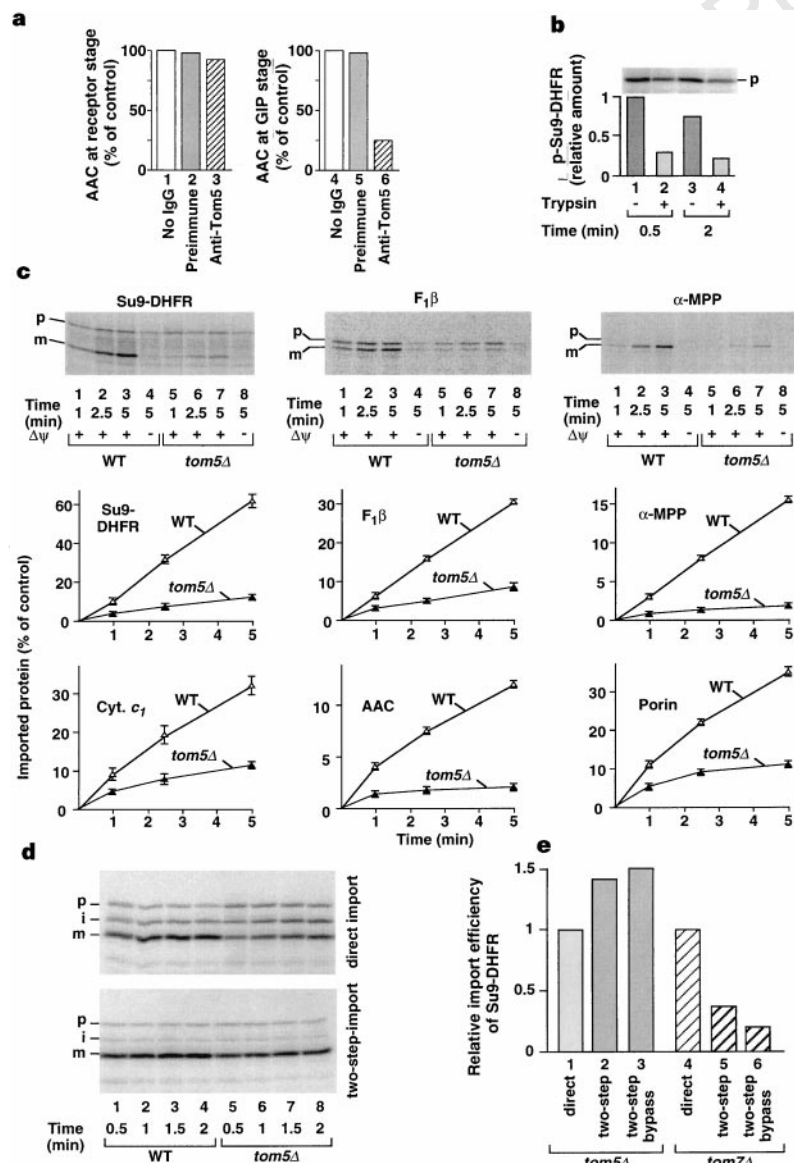


Figure 4 Tom5 functions between receptors and GIP. **a**, IgG (pre-immune or anti-Tom5N) were prebound to wild-type (WT) mitochondria. ADP/ATP carrier was accumulated at receptor sites (samples 1–3) or GIP (samples 4–6; protease treatment after the accumulation¹²). **b**, The *tom5Δ* mitochondria were pretreated with trypsin or mock-treated. After reisolation, mitochondria were incubated with pSu9-DHFR. **c**, Bypass import of preproteins is strongly impaired by lack of Tom5. Mitochondria were pretreated with trypsin. Preproteins were imported as described for Fig. 2b. The amount of import into non-trypsinized WT mitochondria after 5 min was set to 100%. **d**, Two-step import. Su9-DHFR was bound to de-energized mitochondria. Mitochondria were reisolated and a membrane potential ($\Delta\psi$) was generated for the times indicated. For direct import, Su9-DHFR was added in the second incubation (presence of membrane potential). **e**, Direct and two-step import were performed into WT, *tom5Δ* and *tom7Δ* mitochondria for 1 min, and the import ratios between mutant and WT were determined¹¹ (bypass, mitochondria pretreated with trypsin). The relative import efficiency of direct import into the mutant mitochondria was set to 1. Precursor (p), intermediate (i) and mature (m) forms of the preproteins are shown; α -MPP, α -subunit of mitochondrial processing peptidase.

The mild pretreatment of mitochondria with trypsin ($5\text{--}7.5\text{ }\mu\text{g ml}^{-1}$) removes the cytosolic domains of the surface receptors Tom20–Tom22 and Tom70–Tom37, but leaves Tom40 intact. Preproteins can still be imported into mitochondria, although with a reduced efficiency¹⁹; this ‘bypass import’ reflects the direct insertion of preproteins into the GIP^{9,11}. We tested the bypass import of preproteins destined for the distinct subcompartments with wild-type and *tom5* Δ mitochondria, and observed that the receptor-independent import strongly depended on the presence of Tom5 (Fig. 4c). This demonstrates that Tom5 is required for preprotein insertion into the import pore.

When the preprotein was denatured with urea before import, the strong dependence on Tom5 was not relieved (data not shown), indicating that Tom5 is required for import independently of the folding state of a preprotein. We then performed a two-step import of urea-denatured Su9-DHFR. This allows us to define a late function of the outer-membrane translocase in preprotein translocation. The unfolded preprotein is accumulated at the outer membrane in the absence of a membrane potential, and is subsequently imported by regeneration of a membrane potential¹¹. Tom7 and the intermembrane space domain of Tom22 have been found to be required for this productive accumulation of preproteins at the *trans* site of the outer membrane^{4,11,20} (M. Moczko, U.B., M.K., A.H. and N.P., unpublished data). This is demonstrated by a reduction of two-step import in comparison to direct import into *tom7* Δ mitochondria (Fig. 4e, compare columns 5 and 4). The efficiency of two-step import into *tom5* Δ mitochondria was not decreased in comparison to direct import (Fig. 4d, lanes 5–8; Fig. 4e, columns 2 and 1). To exclude an involvement of the surface receptor domains in the productive accumulation, the two-step import reaction was performed with trypsinized mitochondria¹¹ (Fig. 4e, columns 3 and 6). In contrast to the strong inhibition in *tom7* Δ mitochondria (Fig. 4e, column 6), a lack of Tom5 did not inhibit the productive accumulation of the preprotein at the outer-membrane *trans* site (Fig. 4e, column 3). Thus Tom5 is not required at this late stage of preprotein transport across the outer membrane; it functions between surface receptors and the import pore.

With *tom5* Δ mitochondria the N-terminal segment of Tom40, which is exposed to the cytosolic side, was cleaved off by trypsin ($>15\text{ }\mu\text{g ml}^{-1}$), whereas in wild-type mitochondria Tom40 was protected against tryptic attack (data not shown). This finding suggests that Tom5 is closely associated with the N terminus of Tom40 and, because of its own trypsin resistance (Fig. 1d, it protects Tom40 from tryptic digest).

To assess the *in vivo* relevance of Tom5 in cooperation with the other Tom proteins, two genetic approaches were used. First, overexpression of each of the essential Tom proteins, Tom40 and Tom22, did not suppress the defect caused by deletion of *TOM5*, indicating that Tom5 has a function different from that of Tom22 and Tom40. Second, synthetic lethality is observed when mutations or deletions of two genes together cause lethality, whereas cells carrying the mutation or deletion of either gene alone are viable^{11,12,21}. Spores containing double deletions, *tom5* Δ combined with every viable deletion of the other *TOM* genes (*tom6* Δ , *tom7* Δ , *tom20* Δ , *tom37* Δ and *tom70* Δ), were inviable, demonstrating synthetic lethality. Only the double deletion *tom5* Δ *tom72* Δ was viable (Tom72 is a Tom70-homologue of unknown function that loosely assembles with the Tom machinery²). These genetic results are indicative of a functional interaction between Tom5 and most other Tom proteins.

Tom5 is required for the import of preproteins into mitochondria. The Tom proteins identified so far function as receptors, import pore components or modulators of translocase dynamics, but Tom5 provides a new functional element. It represents a central link between the initial recognition event of preproteins (Tom70–Tom37 and/or Tom20–Tom22) and the passage across the membrane (Tom40). By studying a preprotein targeted through Tom70, Tom22 had been proposed to represent the link between import

receptors and GIP²², yet it was found that Tom22 is an essential subunit of the import receptor Tom22–Tom20 (refs 4–6). Preproteins initially recognized by Tom70–Tom37 are transferred to the second receptor subcomplex Tom22–Tom20 before transfer into GIP^{23,24}, explaining the observation of an action of Tom22 between the receptor Tom70 and GIP. Our results show that, after convergence of the targeting pathways at Tom20–Tom22, the preproteins are transferred to Tom5. The negative net charge of its cytosolic segment should facilitate interaction with positively charged targeting sequences. We propose that the cytosolic segment of Tom5 helps to insert preproteins into the GIP. Tom5 thus functions as a mediator between receptors and GIP, and may also represent an early component of the translocation machinery (that is, the GIP) itself. Small subunits have also been identified in other preprotein translocases, such as Sss1/Sec61- γ of the endoplasmic reticulum and SecE in the bacterial export system^{25–27}. Tom5 is the smallest of the known cellular preprotein translocase components, and is the only small component ($<10\text{K}$) for which evidence of a direct interaction with preproteins exists (the 11K protein Tim11 of the mitochondrial inner membrane has also been crosslinked to a preprotein in transit²⁸). The work on Tom6 and Tom7 supported the view of an indirect role of small subunits in preprotein translocases, but the findings made for Tom5 should stimulate investigations on a direct involvement of small subunits in preprotein transport with other translocases as well. □

Methods

Cloning and disruption of *TOM5*. A 1,200-base pair genomic fragment of *TOM5* was amplified by PCR, digested with *EcoRI* and *PstI*, and cloned into pGEM4Z. Inverse PCR mutagenesis was used to delete the *TOM5* ORF and to introduce a *SmaI* and *Sall* site. After digestion with *SmaI* and *Sall*, the marker *HIS3* was introduced. The plasmid was linearized with *EcoRI* and used to transform the diploid yeast strain YPH501 (ref. 11). His³⁺ colonies were selected and allowed to sporulate.

Isolation and characterization of mitochondria and translocase. Isolation of *S. cerevisiae* mitochondria, the outer-membrane translocase (coprecipitation with antibodies directed against Tom40 from digitonin-lysed mitochondria), and outer-membrane vesicles were performed as described^{11,12}. Antisera against an oligopeptide corresponding to the 12 N-terminal residues of Tom5 were prepared by coupling the peptide to keyhole limpet haemocyanin and injection into rabbits. The amino-acid sequence analysis of Tom5 was performed as described^{5,11,29}. For analysis of the Tom complex by two-dimensional gel analysis, 100 μg of wild-type or *tom5* Δ mitochondria were lysed in 1% digitonin-containing buffer and, for the first dimension, separated by blue native electrophoresis^{14,30}. Individual lanes were excised and analysed by urea-SDS-PAGE as second dimension, and immunoblotted onto nitrocellulose. Mitochondria (100 μg mitochondrial protein) were treated at pH 11.5 with 100 mM Na_2CO_3 and separated into supernatant and pellet, followed by immunodecoration¹¹.

Protein import and crosslinking. Preproteins were synthesized in rabbit reticulocyte lysates in the presence of [³⁵S]methionine/[³⁵S]cysteine¹². Import reactions into isolated mitochondria were performed at 25 °C as described¹¹. To allow direct comparison of import reactions, the amounts of preproteins and mitochondria were the same in parallel reactions, and quantifications were made from the linear import ranges¹². The amounts of imported proteins were determined by treatment of mitochondria with proteinase K after the import reaction, followed by separation on SDS-PAGE and digital autoradiography. Surface domains of receptors were removed by a treatment of mitochondria before the import reaction with $5\text{--}7.5\text{ }\mu\text{g ml}^{-1}$ trypsin¹¹. IgG (from preimmune serum or anti-Tom5N serum) was isolated and incubated with isolated mitochondria (150 μg IgG per sample) as described¹². For accumulation of translocation intermediates, ATP was depleted (by apyrase) and/or the membrane potential was dissipated¹². For two-step import¹¹, the membrane potential of mitochondria was dissipated by valinomycin in the presence of potassium, and the urea-denatured preprotein Su9-DHFR was added. Mitochondria were reisolated and a membrane potential was re-established by resuspension in buffer with valinomycin, but without potassium. For direct

import, the precursor was added at the second incubation step. For cross-linking, mitochondria containing accumulated ^{35}S -labelled preproteins were re-isolated through a sucrose cushion^{5,12}, washed and incubated with EGS. After quenching of the crosslinker with Tris buffer, the proteins were precipitated with trichloroacetic acid. The samples were either analysed by SDS-PAGE or lysed in SDS-containing buffer and immunoprecipitated in Triton X-100-containing buffer¹². Standard procedures were used for immunoblotting onto nitrocellulose. Bands were detected by the Enhanced Chemiluminescence System (Amersham). Membrane potential was assessed using DiSC₃(5) (Molecular Probes)^{15,16}.

Received 24 February; accepted 7 May 1997.

- Schatz, G. & Dobberstein, B. Common principles of protein translocation across membranes. *Science* **271**, 1519–1526 (1996).
- Pfanner, N. & Meijer, M. Mitochondrial biogenesis: the Tom and Tim machine. *Curr. Biol.* **7**, R100–R103 (1997).
- Lill, R. & Neupert, W. Mechanisms of protein import across the mitochondrial outer membrane. *Trends Cell Biol.* **6**, 56–61 (1996).
- Hönliger, A., Junne, T., Schatz, G. & Lithgow, T. Acidic receptor domains on both sides of the outer membrane mediate translocation of precursor proteins into yeast mitochondria. *EMBO J.* **14**, 6318–6326 (1995).
- Hönliger, A. *et al.* The mitochondrial receptor complex: Mom22 is essential for cell viability and directly interacts with preproteins. *Mol. Cell. Biol.* **15**, 3382–3389 (1995).
- Mayer, A., Nargang, E. E., Neupert, W. & Lill, R. MOM22 is a receptor for mitochondrial targeting sequences and cooperates with MOM19. *EMBO J.* **14**, 4204–4211 (1995).
- Gratzner, S. *et al.* Mas37p, a novel receptor subunit for protein import into mitochondria. *J. Cell Biol.* **129**, 25–34 (1995).
- Vestweber, D., Brunner, J., Baker, A. & Schatz, G. A 42K outer-membrane protein is a component of the yeast mitochondrial protein import site. *Nature* **341**, 205–209 (1989).
- Kiebler, M. *et al.* Identification of a mitochondrial receptor complex required for recognition and membrane insertion of precursor proteins. *Nature* **348**, 610–616 (1990).
- Alconada, A., Kübrich, M., Moczek, M., Hönliger, A. & Pfanner, N. The mitochondrial receptor complex: the small subunit Mom8b/isp6 supports association of receptors with the general insertion pore and transfer of preproteins. *Mol. Cell. Biol.* **15**, 6196–6205 (1995).
- Hönliger, A. *et al.* Tom7 modulates the dynamics of the mitochondrial outer membrane translocase and plays a pathway-related role in protein import. *EMBO J.* **15**, 2125–2137 (1996).
- Alconada, A., Gärtner, E., Hönliger, A., Kübrich, M. & Pfanner, N. Mitochondrial receptor complex from *Neurospora crassa* and *Saccharomyces cerevisiae*. *Meth. Enzymol.* **260**, 263–286 (1995).
- Steger, H. F. *et al.* Import of ADP/ATP carrier into mitochondria: two receptors act in parallel. *J. Cell Biol.* **111**, 2353–2363 (1990).
- Dekker, P. J. T., Müller, H., Rassow, J. & Pfanner, N. Characterization of the preprotein translocase of the outer mitochondrial membrane by blue native electrophoresis. *Biol. Chem.* **377**, 535–538 (1996).
- Eilers, M., Oppliger, W. & Schatz, G. Both ATP and an energized inner membrane are required to import a purified precursor protein into mitochondria. *EMBO J.* **6**, 1073–1077 (1987).
- Gärtner, F. *et al.* Mitochondrial import of subunit Va of cytochrome c oxidase characterized with yeast mutants: independence from receptors, but requirement for matrix hsp70 translocase function. *J. Biol. Chem.* **270**, 3788–3795 (1995).
- Pfanner, N., Tropschug, M. & Neupert, W. Mitochondrial protein import: nucleoside triphosphates are involved in conferring import-competence to precursors. *Cell* **49**, 815–823 (1987).
- Söllner, T. *et al.* Mapping of the protein import machinery in the mitochondrial outer membrane by crosslinking of translocation intermediates. *Nature* **355**, 84–87 (1992).
- Pfarrer, R., Pfanner, N. & Neupert, W. Mitochondrial protein import: bypass of proteinaceous surface receptors can occur with low specificity and efficiency. *J. Biol. Chem.* **264**, 34–39 (1989).
- Mayer, A., Neupert, W. & Lill, R. Mitochondrial protein import: reversible binding of the presequence at the *trans* side of the outer membrane drives partial translocation and unfolding. *Cell* **80**, 127–137 (1995).
- Huffaker, T. C., Hoyt, M. A. & Botstein, D. Genetic analysis of the yeast cytoskeleton. *Annu. Rev. Genet.* **21**, 259–284 (1987).
- Kiebler, M. *et al.* The mitochondrial receptor complex: a central role of MOM22 in mediating preprotein transfer from receptors to the general insertion pore. *Cell* **74**, 483–492 (1993).
- Haucke, V., Horst, M., Schatz, G. & Lithgow, T. The Mas20p and Mas70p subunits of the protein import receptor of yeast mitochondria interact via the tetrapeptide repeat motif in Mas20p: evidence for a single hetero-oligomeric receptor. *EMBO J.* **15**, 1231–1237 (1996).
- Mihara, K. & Omura, T. Cytoplasmic chaperones in precursor targeting to mitochondria: the role of MSF and hsp70. *Trends Cell Biol.* **6**, 104–108 (1996).
- Brundage, L., Fimmel, C. J., Mizushima, S. & Wickner, W. SecY, SecE, and Band 1 form the membrane-embedded domain of *Escherichia coli* preprotein translocase. *J. Biol. Chem.* **267**, 4166–4170 (1992).
- Esnault, Y., Feldheim, G., Blondel, M. O., Schekman, R. & Képès, F. SSS1 encodes a stabilizing component of the Sec61 subcomplex of the yeast protein translocation apparatus. *J. Biol. Chem.* **269**, 27478–27485 (1994).
- Rapoport, T. A., Junnickel, B. & Kutay, U. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303 (1996).
- Tokatlidis, K. *et al.* Translocation arrest of an intramitochondrial sorting signal next to Tim11 at the inner-membrane import site. *Nature* **384**, 585–588 (1996).
- Eckerskorn, C., Mewes, W., Goretzki, H. & Lottspeich, F. A new siliconized-glass fiber as support for protein-chemical analysis of electrophoretically separated proteins. *Eur. J. Biochem.* **176**, 509–519 (1988).
- Schägger, H., Cramer, W. A. & von Jagow, G. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Analyt. Biochem.* **217**, 220–230 (1994).

Acknowledgements. We thank M. Ryan for comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 388 Freiburg, the Fonds der Chemischen Industrie (N.P.) and a long-term fellowship from the Human Frontier Science Program (P.J.T.D.).

Correspondence and requests for materials should be addressed to N.P. (e-mail: pfanner@ruf.uni-freiburg.de). The sequence of Tom5 has been deposited in the Swiss-Prot protein sequence database, accession no. P80967.

X-chromosome-counting mechanisms that determine nematode sex

Monique Nicoll, Chantal C. Akerib & Barbara J. Meyer

Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720-3204, USA

Sex is determined in *Caenorhabditis elegans* by an X-chromosome-counting mechanism that reliably distinguishes the twofold difference in X-chromosome dose between males (1X) and hermaphrodites (2X)^{1,2}. This small quantitative difference is translated into the 'on/off' response of the target gene, *xol-1*, a switch that specifies the male fate when active and the hermaphrodite fate when inactive³. Specific regions of X contain counted signal elements whose combined dose sets the activity of *xol-1* (ref. 4). Here we ascribe the dose effects of one region to a discrete, protein-encoding gene, *fox-1*. We demonstrate that the dose-sensitive signal elements on chromosome X control *xol-1* through two different molecular mechanisms. One involves the transcriptional repression of *xol-1* in XX animals. The other uses the putative RNA-binding protein encoded by *fox-1* to reduce the level of *xol-1* protein. These two mechanisms of repression act together to ensure the fidelity of the X-chromosome counting process.

The control of sex by the number of X chromosomes was the first genetic regulatory mechanism to be described⁵. In the eighty-one years since its discovery, the mechanism by which chromosomes are counted to determine sex has been established only for *Drosophila*⁶. In *C. elegans*, the *xol-1* gene interprets X-chromosome number^{3,4}. Thus, the question of how the worm counts its X chromosomes has been reduced to the problem of identifying the elements that control *xol-1* activity and the mechanism by which they do so. Sex-specific regulation of *xol-1* is reflected in the tenfold-higher level of *xol-1* transcripts in males than in hermaphrodites³. Setting *xol-1* expression is essential for viability as well as sex determination because *xol-1* also controls X-chromosome dosage compensation⁷. Dosage compensation equalizes X-chromosome expression between the sexes by halving the level of transcripts made from each X chromosome in hermaphrodites^{6,8}. Inappropriate activation of *xol-1* in XX animals prevents the reduction of X expression and thereby causes hermaphrodite-specific lethality³. Repression of *xol-1* expression in XO animals causes inappropriate reduction of X expression and hence male-specific lethality^{3,7}.

To understand how the X-chromosome signal elements regulate *xol-1*, we first investigated whether the sex-specific difference in *xol-1* transcript levels results from transcriptional regulation. The activity of a *lacZ* reporter gene under the transcriptional control of the *xol-1* promoter (Fig. 1g) was assayed in males and hermaphrodites. We observed that XO animals carrying an integrated array (*yIs33*) containing multiple copies of the *Pxol-1::lacZ* reporter transgene had high β -galactosidase activity (Fig. 1c) and that array-bearing XX animals had low activity (Fig. 1b), indicating that transcriptional regulation is central to the sex-specific expression of *xol-1*.

Our previous genetic analysis demonstrated that X-chromosome dose is communicated to *xol-1* by elements in at least three distinct chromosomal regions⁴ (Fig. 1a). Duplication of all three regions causes >99% XO-specific lethality⁴. This lethality is suppressed by constitutive expression of *xol-1*. Deletion of one copy of all three regions in XX animals causes lethality and severe dosage compensation defects that are suppressed by *xol-1* null mutations⁴. To understand how individual components of the signal regulate *xol-1*, we asked whether deletions that eliminate X signal elements affect

the transcription of the *Pxol-1::lacZ* reporter gene. XX embryos from hermaphrodites heterozygous for a deletion that removed the signal element(s) in region 1 (*yDf19*) activated transcription of the *Pxol-1::lacZ* reporter, indicating that region 1 mediates transcriptional control of *xol-1* (Fig. 1d). In contrast, XX embryos from hermaphrodites heterozygous for a deletion that eliminated the signal elements in regions 2 and 3 (*yDf20*) failed to activate the reporter (Fig. 1e). This difference in response cannot be explained by a difference in the relative contribution of the individual X signal elements to the total X signal, because a twofold increase in the dose of regions 2 and 3 causes greater XO-specific lethality (62%) than a

corresponding increase in the dose of region 1 (0%) (Table 1, and ref. 4). Because this lethality is suppressed by constitutive expression of *xol-1*, it is unlikely that the signal elements in regions 2 and 3 regulate a target other than *xol-1*⁴. Thus, the expression studies suggest that the X components of the signal control sex determination and dosage compensation using a combination of two mechanisms, only one of which involves transcriptional repression of *xol-1* in XX animals.

To identify specific X-chromosome signal elements and thereby elucidate the mechanism of X-chromosome counting, we conducted a genetic screen for X-linked suppressors of the XO-specific lethality caused by duplication of the three signal element regions within *mnDp66* and *yDp14* (Fig. 1a, and ref. 4). Based on previous experiments that showed synergism among these signal element regions⁴, we anticipated that loss-of-function mutations in an X signal element would suppress the male lethality by lowering the total dose of signal elements in the duplication-bearing XO animals. From 24,000 mutagenized X chromosomes, we identified two potential signal-element mutations, *y303* and *y304*, and one deficiency that removes region 1. Both *y303* and *y304* were genetically mapped to a small interval covered by *yDp14* that corresponds to region 3 (Fig. 1a).

The genetic properties of the *y303* and *y304* mutations indicate that they define X signal element(s) in region 3. Both *y303* and *y304* are dose-dependent suppressors of the XO-specific lethality caused by duplications that increase the number of X signal elements. The extent of suppression is dependent on the number of wild-type copies of the element identified by *y303* and *y304*. Less than 1% of duplication-bearing XO animals are viable if they have twice the male dose of the elements in regions 1–3 (either one copy of *yDp13* or *mnDp66/yDp14*) (Table 1). In contrast, 100% of these duplication-bearing XO animals are viable if they also carry either *y303* or *y304* (Table 1). Suppression is less effective if more signal elements in region 3 are duplicated. Only 13% of *y303* males and 8.4% of *y304* males are viable if they have two copies of *yDp14*, a dose of signal elements in regions 2 and 3 (Table 1) that would otherwise kill 100% of XO animals.

In addition, *y303* and *y304* cause a synergistic increase in the dosage compensation defects of XX animals that carry deletions of other X signal elements. All XX animals homozygous for either *y303* or *y304* are viable and appear to be completely wild-type (Fig. 2a,b),

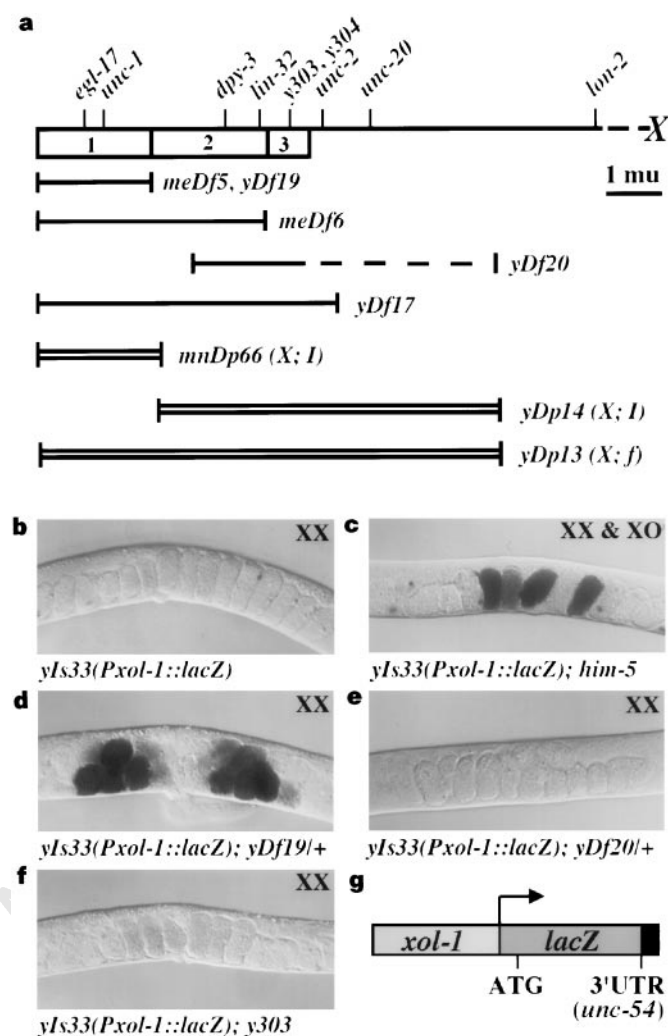


Figure 1 *xol-1* is transcriptionally regulated. **a**, Genetic map of the X chromosome left end shows three dose-sensitive regions (boxed) that contain counted components of the sex-determination signal. Regional boundaries are defined by deficiency (single line) and duplication (double line) endpoints⁴. **b–f**, Photomicrographs of embryo-laden midsections of hermaphrodites histochemically stained for β -galactosidase activity. Each hermaphrodite carries *yIs33*, an integrated array with multiple copies of the *Pxol-1::lacZ* transcriptional reporter gene (**g**). This reporter includes 2.8 kb of *xol-1* genomic 5' regulatory sequences that direct *lacZ* expression. Comparison of β -galactosidase activity in XX embryos (**b**) with that in XO embryos (**c**) shows the transcriptional regulation of *xol-1*. The number of embryos showing intense β -galactosidase expression corresponds to the number of males expected from *him-5* mothers, namely 33% instead of the normal 0.2%. **d**, XX embryos deficient for region 1 (*yDf19*) have high β -galactosidase activity, indicating that signal elements in this region direct the transcriptional repression of *xol-1*. XX embryos deficient for either *fox-1* (**f**) or elements in regions 2 and 3 (*yDf20*) (**e**) have low β -galactosidase activity, suggesting that these signal elements are not transcriptional regulators of *xol-1*.

Table 1 *y303* and *y304* suppress XO-specific lethality caused by duplication of X-chromosome signal elements

Signal elements affected*	Genotype	Male viability (%)	n†
Dp(1) and Dp(2,3)‡	<i>mnDp66/yDp14</i> ; +	1	660
Dp(1) and Dp(2,3) and <i>y303</i> ‡	<i>mnDp66/yDp14</i> ; <i>y303</i>	109	1,031
Dp(1) and Dp(2,3) and <i>y304</i> ‡	<i>mnDp66/yDp14</i> ; <i>y304</i>	96	1,450
Dp(1,2,3)§	+; <i>yDp13</i>	0.4	266
Dp(1,2,3) and <i>y303</i> §	<i>y303</i> ; <i>yDp13</i>	90	1,001
Dp(1,2,3) and <i>y304</i> §	<i>y304</i> ; <i>yDp13</i>	94	747
Dp(2,3) and Dp(2,3)	<i>yDp14/yDp14</i> ; +	0	>1,000
Dp(2,3) and Dp(2,3) and <i>y303</i>	<i>yDp14/yDp14</i> ; <i>y303</i>	13	507
Dp(2,3) and Dp(2,3) and <i>y304</i>	<i>yDp14/yDp14</i> ; <i>y304</i>	8.4	745
Dp(2,3)¶	<i>yDp14</i> /+	38	814
Dp(1)¶	<i>mnDp66</i> /+	102	1,874

* Dps are X-chromosome duplications containing one or more of the three signal element regions (Fig. 1a), indicated in parentheses. *yDp13* is a free duplication; *yDp14* and *mnDp66* are attached to chromosome I.

† n, Total number of animals counted.

‡ *yDp14*/+; *him-8*; *unc-2* males were crossed with hermaphrodites of genotype *mnDp66*; *unc-1* *dpy-3* or *mnDp66*; *unc-1* *dpy-3* *y303* or *mnDp66*; *unc-1* *dpy-3* *y304*, and the per cent male viability was calculated as (total non-Dpy non-Unc males/one-half the total non-Dpy non-Unc hermaphrodites) $\times$ 100.

§ *meDf6*; *yDp13* males were crossed with hermaphrodites of genotype *dpy-3* *unc-2* or *dpy-3* *y303* *unc-2* or *dpy-3* *y304* *unc-2*, and the per cent male viability was calculated as (total non-Dpy non-Unc males/total non-Dpy non-Unc hermaphrodites) $\times$ 100.

|| Total live self progeny were counted from several individual hermaphrodites of genotype *yDp14/yDp14*; *him-8*; *unc-2* or *yDp14/yDp14*; *him-8*; *dpy-3* *y303* *unc-2* or *yDp14/yDp14*; *him-8*; *dpy-3* *y304* *unc-2*. Per cent male viability was calculated as (total males)/(0.37 $\times$ total expected progeny). Total expected progeny was calculated as (total hermaphrodites/0.63). These calculations are based on the fact that 37% of the self progeny of *him-8*(e1489) hermaphrodites are XO (ref. 14).

¶ Data from ref. 4.

as are XX animals heterozygous for *meDf6*, a deletion that removes the signal elements in regions 1 and 2. However, the combination of lesions in the *y303/meDf6* trans-heterozygous XX animals causes pronounced dosage-compensation defects (Fig. 2c). These dosage-compensation defects are comparable to those caused by deletions that remove all three signal element regions, either *yDf17* or the heterozygous combination of *yDf20/meDf5* (Fig. 2d). This genetic analysis reveals an important role for the locus defined by *y303* in counting X chromosomes and suggests that this locus can account for most, if not all of the X signal activity in region 3.

y303 and *y304* map to a region of X previously identified as causing XO-specific lethality and feminization when present in multiple copies⁹. The genetic locus responsible for the male lethal activity was not defined, but the activity was named *fox-1* (for 'feminizing locus on X') and was narrowed to a 30-kilobase (kb) segment of DNA⁹. A 1.8-kb embryonically expressed complementary DNA from within this segment was predicted to encode a 415-amino-acid protein with an RNP motif⁹, suggesting that the protein binds to RNA¹⁰. DNA sequence analysis of this RNP gene from our mutant strains revealed that the *y303* mutation causes an ochre translational stop at codon 37 and is therefore likely to cause complete loss of function (Fig. 2e). The *y304* allele causes a proline-to-serine change in the RNP motif at amino acid 161 of the same protein (Fig. 2e). This proline is predicted to be in a small loop region, between the β 1 sheet and the α 1 helix, and is conserved among many of the RNP proteins. This molecular information, together with our genetic analysis of *y303* and *y304*, establishes that the RNP gene, which we designate *fox-1*, is a *bona fide* X-chromosome sex-determination signal element. Moreover, because *y303* XX and XO animals are fully viable and fertile, and *y303* is probably a null allele, *fox-1* appears not to be an essential gene.

To learn how the FOX-1 protein functions in the X-chromosome counting process, polyclonal antibodies were raised against the entire 415-amino-acid protein and were used to determine the protein's location in wild-type and *fox-1*-mutant embryos. We expected a signal element to be expressed by the 28-cell stage and through gastrulation (400-cell stage) because *xol-1* must be active during that early period to trigger male development and must be repressed during that period to permit hermaphrodite development³. We observed very early nuclear staining in the cells of wild-type XX embryos beginning at the 8–16 cell stage of development; the staining intensity peaked by the 100-cell stage and was absent by the end of embryonic cell proliferation around the 550-cell stage (Fig. 2f). XX embryos with the *fox-1(y303)* nonsense mutation lacked staining, establishing the specificity of the antibody and confirming the classification of *fox-1(y303)* as a null allele (Fig. 2f). The staining results indicate that FOX-1 is present during an appropriate window of time to act as a component of the X signal and to downregulate *xol-1* in hermaphrodites.

fox-1 does not seem to regulate the transcription of *xol-1*. Transcriptional repression of *xol-1*, as monitored by the *xol-1::lacZ* transcriptional fusion (*yIs33*), was robust in XX embryos lacking *fox-1* (Fig. 1f), as we had found for XX embryos lacking the elements in regions 2 and 3 (Fig. 1e). To demonstrate that FOX-1 nonetheless acts on *xol-1* and to address the mechanism of repression, we used a different type of *xol-1* reporter gene, one that contained all the *xol-1* genomic sequences. In these experiments, the coding region of the green fluorescent protein (GFP) was fused in-frame to the first ATG codon of *xol-1*, creating a fully bifunctional GFP::XOL-1 fusion protein (Fig. 3d). Eight independent extrachromosomal arrays carrying multiple copies of this *gfp::xol-1* transgene were established in *xol-1* mutant strains. In each case, the array suppressed the XO-specific

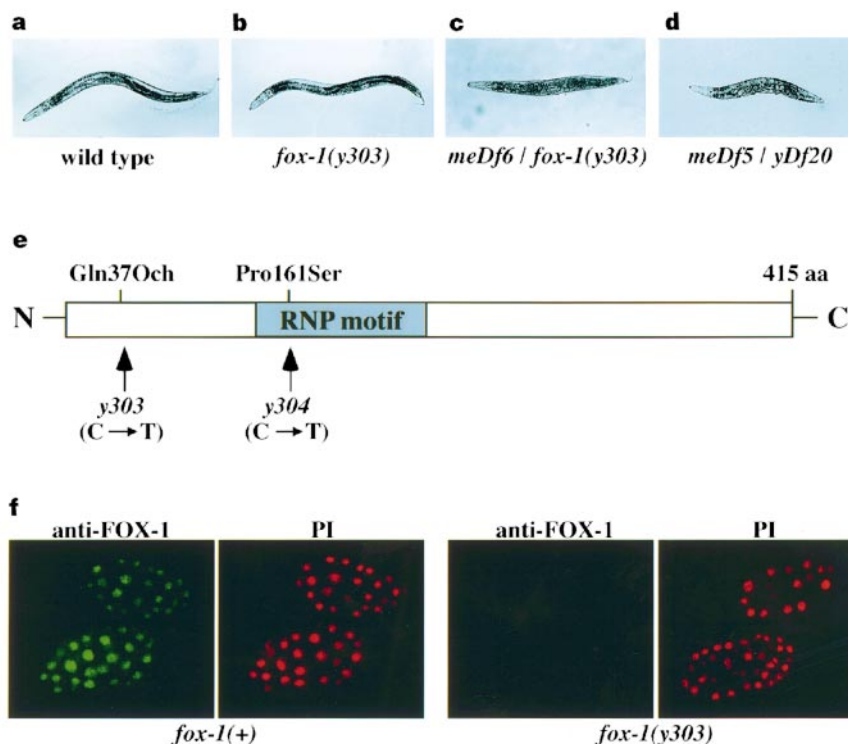


Figure 2 The putative RNA binding protein, FOX-1, is a signal element. **a–d**, Bright-field photomicrographs comparing wild-type XX animals (**a**) to XX animals with X chromosome signal element mutations. **b**, *fox-1* mutants appear wild type. **c**, Loss of one copy of *fox-1* and one copy of elements in regions 1 and 2 causes synergistic dosage compensation defects (dumpy and egg-laying-defective phenotypes) similar to those caused by loss of one copy of signal elements in

regions 1–3 (**d**). **e**, Signal element mutations *y303* and *y304* cause changes in the putative RNA binding protein, FOX-1. **f**, Confocal images of wild-type and *fox-1(y303)* embryos co-stained with FOX-1 antibody (green) and the DNA intercalating dye, propidium iodide (PI), (red). FOX-1 is nuclear localized in wild-type embryos and absent in *y303* embryos.

lethality caused by the *xol-1* mutation, indicating that the fusion protein has XOL-1 activity. All the arrays expressed high levels of GFP in XO (Fig. 3b) but not XX (Fig. 3a) animals. Moreover, none of these arrays caused XX-specific lethality or dosage compensation defects. Thus, these *gfp::xol-1* reporter arrays showed proper sex-specific regulation. They were crossed into strains that produce high levels of FOX-1 from an integrated array (*yls44*) containing multiple copies of the *fox-1* gene. The level of FOX-1 produced by *yls44* causes complete XO-specific lethality in otherwise wild-type strains. In *yls44* strains with the *gfp::xol-1* arrays, no GFP expression was detected (Fig. 3c) and all the XO animals died. These results indicate that high doses of FOX-1 drastically reduce the levels of both the GFP::XOL-1 protein and the endogenous XOL-1 protein, revealing that FOX-1 acts through *xol-1*.

In contrast to the eight *gfp::xol-1* arrays just described, which caused no XX-specific lethality, two other *gfp::xol-1* arrays expressed GFP::XOL-1 at a level that caused complete XX-specific lethality in strains with or without excess FOX-1. These two arrays also suppressed the male-lethal effect of excess FOX-1, suggesting that the FOX-1 activity had been titrated by *xol-1* sequences or protein. Such an interpretation is consistent with the fact that *xol-1* acts downstream of *fox-1*, and it suggests the possibility that *xol-1* may be the direct target of FOX-1.

Additional evidence for a post-transcriptional role for FOX-1 was provided by a different *xol-1* reporter construct in which the *gfp* coding sequences were fused in frame to a portion of the *xol-1* gene containing the promoter region, the 5'UTR, the first two introns and the first 89 codons (Fig. 3h). This fusion, called *Pxol-1::gfp*, does not provide *xol-1*(+) activity. Six independent extrachromosomal arrays carrying multiple copies of this reporter gene cause strong GFP staining in XO (Fig. 3f) but not XX (Fig. 3e) animals,

mimicking the sex-specific expression of *xol-1*. Excess FOX-1 produced from *yls44* failed to repress this translational *Pxol-1::gfp* fusion, indicating that the *fox-1* target sequences are not present in this construct (Fig. 3g). These results, together with the nuclear localization and the similarity of FOX-1 to RNA-binding proteins, suggest that FOX-1 regulates *xol-1* at a post-transcriptional step, such as splicing, stability, transport or translation of the *xol-1* RNA.

Our results reveal the general molecular strategy by which *C. elegans* amplifies the twofold difference in X chromosome dose to set sexual fate through effects on *xol-1* (Fig. 3i). One class of X signal elements represses the transcription of *xol-1* in XX animals. A second class, including the RNA-binding protein FOX-1, further reduces *xol-1* activity in XX animals through a post-transcriptional mechanism. The absence of obvious defects in XX animals lacking *fox-1* activity does not mean that the post-transcriptional mechanism has little significance in the counting process, because elements in region 2 that participate in this aspect of regulation remain in these mutants. Indeed, a female dose of the combined post-transcriptional elements in regions 2 and 3 reduces the viability of XO animals by half. The significance of the post-transcriptional mechanism is further demonstrated by the synergistic effect of mutations that partially inactivate both mechanisms in XX animals. Hermaphrodites with a male dose of region 1 (transcriptional) or a male dose of regions 2 and 3 (post-transcriptional) are wild-type, but hermaphrodites with a male dose of all three regions exhibit severe dosage compensation defects and death. Thus, the two mechanisms of repression act together to ensure the fidelity of the X chromosome counting process.

We can compare the X-chromosome counting process in *C. elegans* with that in *Drosophila melanogaster*. The fruitfly uses the XX dose of its signal elements to activate transcription of the target

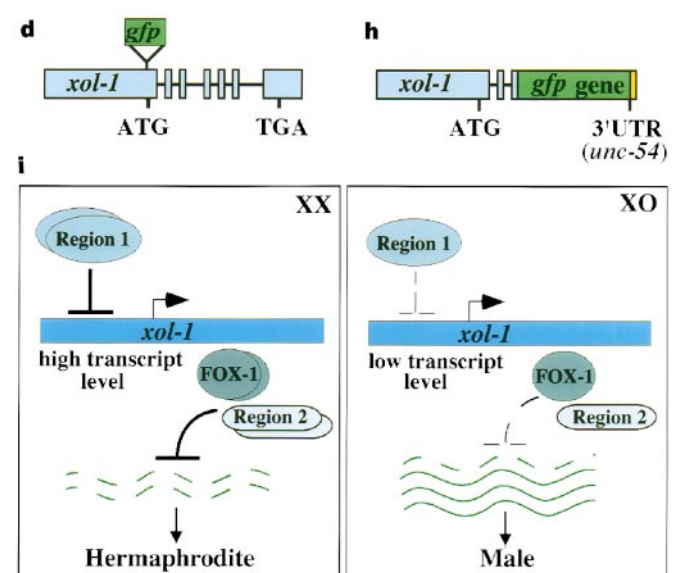
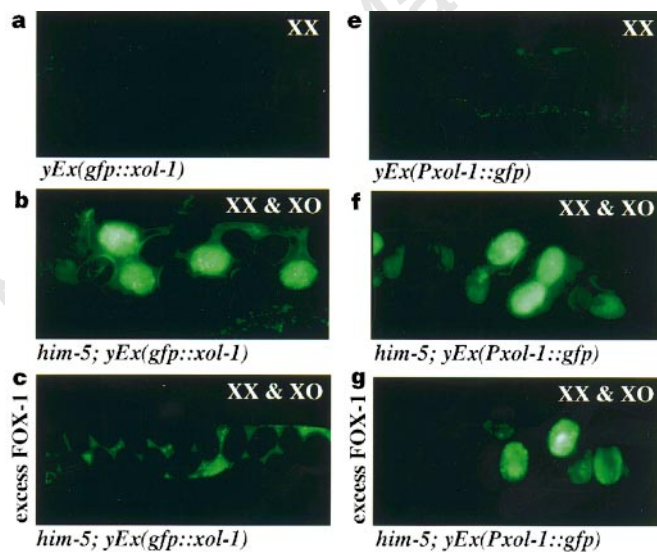


Figure 3 *fox-1* represses *xol-1*. **a–c**, Fluorescence photomicrographs of embryo-laden midsections of hermaphrodites carrying *yEx(gfp::xol-1)*, an extrachromosomal array with multiple copies of the *gfp::xol-1* reporter gene in **d**. In this reporter, the GFP coding sequence was fused in frame to the first ATG of *xol-1*, thus creating a bifunctional GFP::XOL-1 fusion protein. The sex-specific expression of *yEx(gfp::xol-1)* is shown by the absence of GFP fluorescence in XX embryos (**a**), and the presence of GFP fluorescence in XO embryos (**b**). **c**, Excess FOX-1 provided from an integrated *fox-1*(+) array, *yls44*, represses embryonic GFP fluorescence, showing that FOX-1 acts through *xol-1*. **e–g**, Midsections of hermaphrodites carrying *yEx(Pxol-1::gfp)*, an extrachromosomal

array containing multiple copies of the *Pxol-1::gfp* translational reporter gene in **h**. In this reporter, the *gfp* coding region replaces *xol-1* sequences downstream of the 89th codon. **e**, **f**, *Pxol-1::gfp* is sex-specifically expressed. **g**, Excess FOX-1 fails to repress GFP fluorescence, indicating that *Pxol-1::gfp* lacks FOX-1 target sequences. **i**, A model for X chromosome counting by *xol-1*. In XX embryos, *xol-1* is inactivated in response to a hermaphrodite dose of X-chromosome signal elements. Complete repression of *xol-1* activity involves at least two concerted mechanisms, one acting transcriptionally and a second, mediated by the FOX-1 RNA-binding protein, acting post-transcriptionally. In XO embryos, *xol-1* remains active despite the male dose of X-chromosome signal elements.

switch gene, *Sex-lethal*, and thereby initiate female development⁶, whereas *C. elegans* represses expression of its target *xol-1* through transcriptional and post-transcriptional mechanisms. No evidence exists for post-transcriptional regulation operating in the initial X counting process of flies, although they do use a post-transcriptional mechanism to lock in the counting choice⁶. Mammals must also assess their X-chromosome number to achieve dosage compensation through X inactivation. Although the mechanism has not been determined, the evidence available¹¹ implies that it is likely to differ from the chromosome-counting mechanisms in both flies and worms. □

Methods

Genetic mapping. *y303* and *y304* were isolated using a previously described screen⁴. Of 948 embryos counted from *fox-1(y303)* XX hermaphrodites, all developed into fertile adults. *y303* was mapped between *lin-32* and *unc-2* by analysing Unc non-Dpy and Dpy non-Unc recombinant progeny from *dpy-3 + y303 unc-2/+ lin-32 + +* mothers. Of 21 Unc non-Dpy recombinants, 4 recombined between *dpy-3* and *lin-32*, 5 between *lin-32* and *y303*, and 12 between *y303* and *unc-2*. Of 6 Dpy non-Unc recombinants, 2 recombined between *lin-32* and *y303* and 4 between *y303* and *unc-2*. *y304* was similarly mapped. Of 16 Unc non-Dpy recombinants, 4 recombined between *dpy-3* and *lin-32*, 7 between *lin-32* and *y304*, and 5 between *y304* and *unc-2*. Of 17 Dpy non-Unc recombinants, 5 recombined between *lin-32* and *y303*, and 12 between *y304* and *unc-2*. *yDf19* is a γ ray-induced mutation that fails to complement *egl-17* and *unc-1* and does not cause X-chromosome non-disjunction. Of 976 self progeny from *yDf19/unc-1 dpy-3* mothers, none was male. *meDf6* removes *lin-32* but probably not *fox-1*. Fourteen dead embryos from *meDf6/dpy-3 unc-2* mothers amplified a control band and a *fox-1*-specific band in single-embryo polymerase chain reactions (PCRs). *yDf20* is a suppressor of the XO-specific lethality caused by two copies of *yDp14* (I. Carmi and B.J.M., unpublished results) and deletes *dpy-3*, *lin-32* and *fox-1*. Fifteen dead embryos from *yDf20/unc-1 dpy-3* mothers that amplified a control band failed to amplify a *fox-1*-specific band in single-embryo PCR. *yDf17* was isolated as a suppressor of the XO-specific lethality caused by *mnDp66/yDp14*. *yDf17* removes the left end of X, including *unc-2*, and *yDf17/+* XX animals have an Sdc phenotype similar to that of *yDf14/+* XX animals (ref. 4, and I. Carmi and B.J.M., unpublished results). For single-embryo PCRs¹², the primer pairs 5'-CCCATTTCAGATTCAGAGACC and 5'-CGAGTGAACACGAGCTGTAG amplified an internal fragment of *fox-1* (ref. 9). The control primer pairs 5'-CTACTGTCGACAATGTTGGAATCCTC and 5'-GGGATTCTGCAGTTGCAAGATG amplified a fragment of *sdC-3*.

DNA sequence analysis of *y303* and *y304*. The DNA changes caused by *y303* and *y304* were determined by DNA sequence analysis of cDNA generated by reverse-transcribed PCRs of RNA extracted from mutant (*dpy-3 y303* and *dpy-3 y304*) and parental (*mnDp66; unc-1 dpy-3*) strains. *fox-1* was amplified using the primers 5'-CCGCTCGAGATGCAAGCCCTGTACCAACT or 5'-CAGTCGGCGTTTGAATGGATCC with 5'-CGGGATCCCACTCAATACGGAGTAAATCG. The PCR products for each strain were cloned for three independent reactions and were sequenced from three clones. All three sets of sequences from the *y303* strain revealed only a C-to-T transition at nucleotide 109, and from the *y304* strain, a C-to-T transition at nucleotide 481.

***xol-1* reporter genes and transgenic arrays.** The *Pxol::lacZ* transcriptional fusion (pMN21) was made by insertion of the 2,779-bp *EcoRI* fragment from the *xol-1* promoter region into the *SmaI* site of pPD95.03, an intron-rich version of pPD16.01 (ref. 13) (from A. Fire). The *xol-1/lacZ* junction occurs 71 bp upstream of the first ATG in *xol-1*, thus eliminating *xol-1*'s normal 5'UTR, made by the insertion of an SL1 trans-spliced leader 13 nucleotides upstream of that ATG. *yls33* was made by integrating an extrachromosomal array containing pRF4, a plasmid encoding a dominant *rol-6(su1006)* marker, and pMN21 coinjected at 100:40 $\mu\text{g ml}^{-1}$. The *gfp*-tagged *xol-1* genomic fusion (pMN45) (*gfp::xol-1*) was made by inserting a PCR fragment of *gfp* coding sequences (a S65T, I167T variant from Y. Jin) in frame at the first ATG of *xol-1*. The *yEx(gfp::xol-1)* extrachromosomal arrays contain p76-16B, a plasmid that encodes *unc-76(+)*, and pMN45 co-injected at 150:10 $\mu\text{g ml}^{-1}$. The *Pxol-1::gfp* translational fusion (pMN15) was made by replacing the 3,441-bp *SphI*-*Apal* fragment from genomic *xol-1* sequences with the 1,931-bp *SphI*-*Apal* *gfp*

fragment from the expression vector pPD95.67, which encodes the S65C variant of GFP (from A. Fire). The *yEx(Pxol-1::gfp)* extrachromosomal arrays contain p76-16B and pMN15 coinjected at 150:20 $\mu\text{g ml}^{-1}$. More than one thousand embryos of each appropriate genotype were scored in experiments using *xol-1* reporter constructs. The XO-specific expression of transgenic arrays carrying these reporter constructs was shown by the lack of embryonic activity in transgenic strains that produced only wild-type XX embryos and the abundant embryonic activity in transgenic lines that produced wild-type XO embryos because they carried a *him-5* mutation¹⁴. Excess *fox-1* was produced from an integrated array, *yls44*, which causes XO-specific lethality and contains multiple copies of the *fox-1*-containing cosmid R04B3 and pRF4.

FOX-1 antibodies. Rabbit polyclonal antibodies were raised against a bacterially expressed HIS6::FOX-1 fusion protein that included the entire FOX-1 protein. The *fox-1* open reading frame was cloned into the expression vector pRSET-A (Invitrogen), and the HIS6::FOX-1 fusion protein was expressed and purified according to the procedures of Qiagen. Antibodies were affinity-purified using a fusion protein containing the entire FOX-1 protein and the maltose-binding protein (New England Biolabs). Embryos were stained as described¹⁵ using a 1:50 dilution of affinity-purified FOX-1 antibodies.

Received 2 April; accepted 13 May 1997.

1. Nigon, V. Polyplodie expérimentale chez un nematode libre, *Rhabditis elegans* Maupas. *Bull. Biol. Fr. Belg.* **85**, 187–225 (1951).
2. Madl, J. E. & Herman, R. K. Polyplodies and sex determination in *Caenorhabditis elegans*. *Genetics* **93**, 393–402 (1979).
3. Rhind, N. R., Miller, L. M., Kopczynski, J. B. & Meyer, B. J. *xol-1* acts as an early switch in the *C. elegans* male/hermaphrodite decision. *Cell* **80**, 71–82 (1995).
4. Akerib, C. C. & Meyer, B. J. Identification of X chromosome regions in *Caenorhabditis elegans* that contain sex-determination signal elements. *Genetics* **138**, 1105–1125 (1994).
5. Bridges, C. B. Non-disjunction as proof of the chromosome theory of heredity. *Genetics* **1**, 1–52 (1916).
6. Cline, T. W. & Meyer, B. J. *Vive la difference*: males vs females in flies vs worms. *Annu. Rev. Genet.* **30**, 637–702 (1996).
7. Miller, L. M., Plenefisch, J. D., Casson, L. P. & Meyer, B. J. *xol-1*: A gene that controls the male modes of both sex determination and X chromosome dosage compensation in *C. elegans*. *Cell* **55**, 167–183 (1988).
8. Meyer, B. J. & Casson, L. P. *Caenorhabditis elegans* compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. *Cell* **47**, 871–881 (1986).
9. Hodgkin, J., Zellman, J. D. & Albertson, D. G. Identification of a candidate primary sex determination locus, *fox-1* on the X chromosome of *Caenorhabditis elegans*. *Development* **120**, 3681–3689 (1994).
10. Bird, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621 (1994).
11. Willard, H. F. X chromosome inactivation, XIST, and pursuit of the X-inactivation center. *Cell* **86**, 5–7 (1996).
12. Barstead, R. J., Kleiman, L. & Waterston, R. H. Cloning, sequencing and mapping of an α -actinin gene from the nematode *Caenorhabditis elegans*. *Cell Motil. Cytoskel.* **20**, 69–78 (1991).
13. Fire, A., Harrison, S. W. & Dixon, D. A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189–198 (1990).
14. Hodgkin, J., Horvitz, H. R. & Brenner, S. Nondisjunction mutants of *Caenorhabditis elegans*. *Genetics* **91**, 67–94 (1979).
15. Chuang, P.-T., Albertson, D. G. & Meyer, B. J. DPY-27: A chromosome condensation homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**, 459–474 (1994).

Acknowledgements. We thank I. Carmi, T. Cline, J. Rine, H. Dawes for discussions and critical comments on the manuscript. A. Fire for reporter gene vectors and Y. Jin for the *gfp* coding sequences. M.N. is a National Science Foundation predoctoral fellow. C.C.A. was an American Cancer Society postdoctoral fellow. This work was supported by a grant from the National Institutes of Health to B.J.M.

Correspondence and requests for materials should be addressed to B.J.M. (e-mail: meyer@mendel.berkeley.edu).

correction

Structure of the adenylyl cyclase catalytic core

Gongyi Zhang, Yu Liu, Arnold E. Ruoho & James H. Hurley

Nature **386**, 247–253 (1997)

In Fig. 3, the sequence in line 2 from 908 to 943 should read PGEIVHMLNELFGKFDQIAKENEC...MRIKLGDC, and from 1,004 to 1,022 should be NVLCGVIGLQKWQYDVVWSH. An extraneous 19 amino acids appeared at the start of the third line of the last block of sequence. This line should therefore start with AGGRA.... Also, the correct GenBank accession code for bovine type 1 adenylyl cyclase is M25579. □

Oligosaccharide sequencing technology

Recent advances allow any well-founded laboratory to carry out rapid and straightforward characterization of subpicomolar levels of *N*-glycans released from one to five microgram quantities of glycoprotein.

Pauline M. Rudd, Geoffrey R. Guile, Bernhard Küster, David J. Harvey, Ghislain Opdenakker & Raymond A. Dwek

The glycosylation machinery in a cell is available to all proteins that enter the secretory pathway. Two major classes of oligosaccharides may be attached to proteins that contain the appropriate signals in their peptide chains. *N*-linked sugars may be added to the amide side-chain of some Asn residues, which form part of the triplet AsnXaaSer/Thr, while *O*-linked sugars may be added to the hydroxyl side-chain of some Ser or Thr residues. Glycoproteins generally consist of mixtures of glycosylated variants of a single polypeptide (glycoforms). The composition of the glycoform population is influenced by the 3-dimensional structure of the protein¹.

A full understanding of the implications of glycosylation can only be reached when the glycoprotein is viewed as an entity. While peptide sequencing is routinely available to protein chemists there is also a need for a robust, rapid and automated technology for oligosaccharide sequencing at the subpicomolar level, which is applicable to all glycoproteins.

Earlier developments in sequencing

The first predictive approach to oligosaccharide analysis was based on BioGel P4 gel permeation chromatographic separations of glycan pools and exoglycosidase sequencing of individual sugars² using enzyme arrays³. More recently, the higher resolving power of high-performance liquid chromatography (HPLC), first used for analysing sugars in 1975 (refs 4 and 5), has been exploited⁶⁻⁸. The use of 2- and 3-dimensional glycan analysis in which glycan pools are resolved into individual sugars by successive passages through anion-exchange, hydrophobic and hydrophilic interaction HPLC columns was pioneered mainly by Japanese scientists^{9,10}. The glycan structures are assigned by comparing the 2- or 3-D coordinates of the elution positions of the samples with those of known standards. Other separation techniques include capillary electrophoresis¹¹⁻¹² and polyacrylamide gel electrophoresis¹³. Mass spectrometry is an important method for determining oligosaccharide composition. NMR can determine unambiguously the complete structure of an oligosaccharide

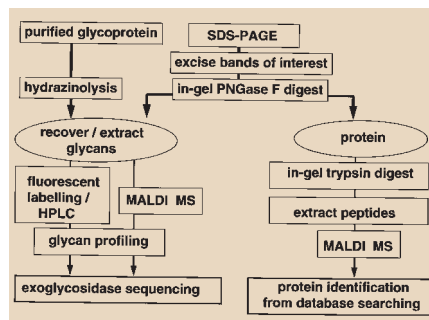


Figure 1 Strategies for glycan analysis and protein identification.

where sufficient material is available. Clearly oligosaccharide analysis is an advancing field to which many groups have contributed²⁻¹⁶.

A strategy for *N*-glycan analysis

The strategy illustrated in Fig. 1 specifically addresses the need for a rapid and robust method of *N*-glycan analysis, which can be applied routinely to microgram levels of glycoproteins with the minimum requirement for specialized equipment or expertise. The steps involved in oligosaccharide analysis include releasing the sugars from the protein, and labelling, separating and analysing the components of the glycan pool.

For samples where the protein is difficult to purify or amounts are limited, the *N*-glycans may be released directly from a band on an SDS-PAGE gel using peptide *N*-glycosidase F (PNGase F)¹⁷.

For detection in HPLC analysis, the non-reducing termini of the glycans are labelled with 2-aminobenzamide (2AB)¹⁸, to allow direct quantitation from the HPLC profiles. In addition, the reaction conditions retain sialic acid residues and the label is compatible with other analytical systems such as matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS).

An amide-silica HPLC column was selected because the system could be optimized to minimize interaction of the column with the fluorescent label. High-resolution separations of both charged and neutral *N*- and *O*-glycans were achieved in a single run. The system may be used preparatively, as well as analytically. A volatile buffer system (acetonitrile and ammonium formate) was selected so that individual glycans could be recovered free of buffer salts. A major advantage of this system is that the elution positions of the glycans may be used

to predict structure, as positive incremental values are obtained for the addition of all monosaccharide residues to basic core oligosaccharides^{19,20}.

Assignments may be confirmed by sequential exoglycosidase digestion using highly specific enzymes. The products of such digestions are oligosaccharide fragments from which the enzyme(s) have removed specific terminal residues. As a result of the high resolution of the HPLC system, a rapid sequencing method has now been developed. In this approach, five aliquots of the entire glycan pool are sequenced simultaneously using a standard panel of enzyme arrays, after which the products are analysed using the same HPLC system. Any unusual structures are highlighted since these will not be digested to a common trimannosyl *N*-linked glycan core by any of the arrays. These structures may be isolated from preparative HPLC runs for further analysis.

Predictive HPLC glycan profiling

A novel, sensitive and reproducible HPLC technology has been developed to resolve subpicomole levels of heterogeneous mixtures of sugars fluorescently labelled with 2-aminobenzamide in their correct molar proportions¹⁸.

N- and *O*-linked oligosaccharides released from glycoproteins are resolved on the basis of differences in hydrophilicity, which reflects arm specificity, linkage and monosaccharide composition. In contrast to earlier technology based on gel filtration², structures of both the neutral and charged components of an entire *N*-glycan pool may be predicted from a single HPLC run (Fig. 2a, b).

The structures of the glycans are predicted from their HPLC elution positions determined in glucose units with reference to a dextran ladder. Assignments are based on the elution positions of known glycans from which incremental values for the addition of monosaccharide residues to oligosaccharide cores have been calculated (Fig. 2b). The quality of the resolution and the ability to predict glycan structures allow the interpretation of the full oligosaccharide profiles of glycoproteins, which may contain 30 or more structures at each glycosylation site (Fig. 3).

A major advantage of all HPLC systems is that they can be used preparatively. This means, for example, that the glycans in individual peaks can be collected, the

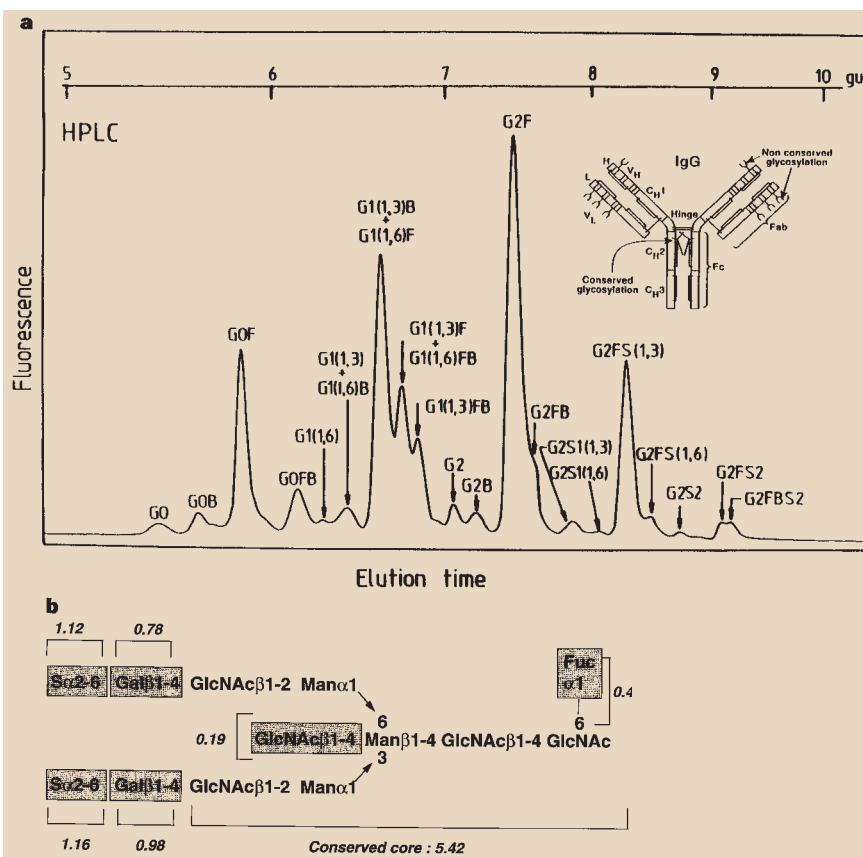


Figure 2a, Normal phase HPLC separation of glycans from normal human serum IgG. Peaks were assigned glucose unit values by comparison with a standard dextran ladder, the elution positions of which are shown at the top of the figure. The inserted diagram of IgG indicates the conserved *N*-glycosylation sites in the Fc and shows variable glycosylation sites in the Fab. (Nomenclature for describing oligosaccharide structures: A(1,2) indicates the number of antennae linked to the trimannosyl core; G(0–2) indicates the number of terminal galactose residues in the structure; F: fucose; B: bisecting *N*-acetylglucosamine (GlcNAc); S: sialic acid; Gal: galactose.) **b**, Complex biantennary sugar from which the set of IgG sugars is derived. The presence or absence of the boxed monosaccharide residues gives rise to the heterogeneity seen in panel **a**. The elution positions depend on the hydrophilic interactions of each oligosaccharide with the HPLC column and are sensitive to the environment of the monosaccharide residues within the glycan structure. This gives the column its fine specificity. For example, the incremental values for the addition of galactose to the α 1,3 arm differs from the addition of the same residue to the α 1,6 arm. Tables of incremental values are given in refs 19,20. These form the basis of the HPLC predictive technique. To illustrate this, incremental values for the addition of the different monosaccharide residues to the trimannosyl core to obtain the fucosylated, bisected, disialylated sugar are shown in this figure.

volatile buffers removed, and the glycans analysed by mass spectrometry or capillary electrophoresis to validate assignments. In cases where unusual or novel structures are indicated, further information can also be obtained by 2-D chromatography. For example, glycan pools containing sulphated sugars isolated from normal phase separations may be further resolved by weak anion exchange chromatography²¹.

The new technology enables a detailed comparison of glycosylation profiles and has been used, for example, to compare IgG glycans from normal and rheumatoid IgG (ref. 22). Predictive HPLC analysis can be used to gain an overview of the entire glycosylation

of a glycoprotein (Fig. 3). Rapid profiling enables recombinant glycoproteins to be monitored during cell culture. Furthermore, the ability to resolve neutral and acidic glycans simultaneously is particularly useful where acidic sugars are likely to play a functional role as, for example, in the brain²³. Current studies are focused on extending the methodology to *O*-glycans which are resolved by the same system.

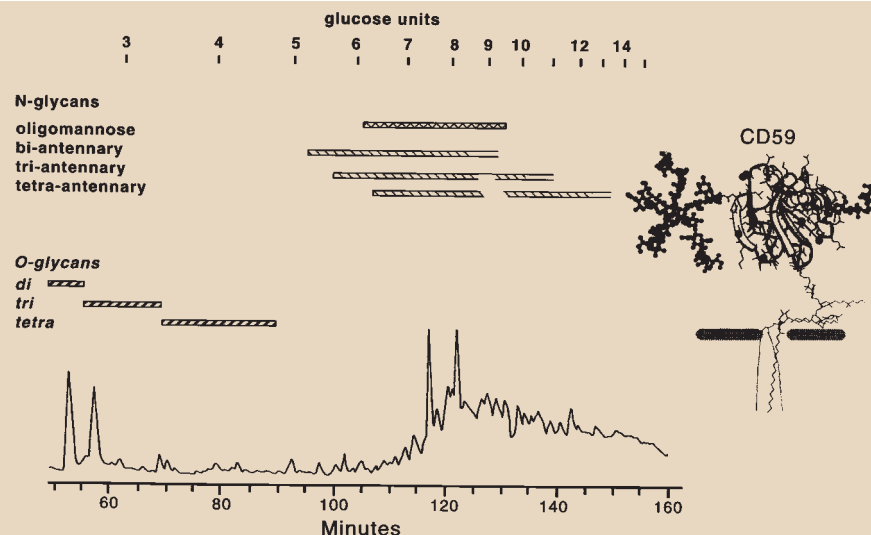
Simultaneous analysis of glycan pools

Further structural information can be obtained rapidly by simultaneously sequencing all of the oligosaccharides in a glycan pool. Enzymatic analysis of oligosaccharides

using highly specific exoglycosidases is a powerful means of determining the sequence and structure of glycan chains. Until recently, this has involved the isolation of single glycans from the glycan pool. It is often impractical and always time consuming to purify each individual sugar to the necessary 80% purity.

The new approach involves the simultaneous digestion of aliquots of a total pool of fluorescently labelled oligosaccharides with a series of multiple enzyme arrays²⁴. The products are analysed by normal phase HPLC or by MALDI MS, and peaks are assigned from a knowledge of the specificity of the enzymes and the incremental values of individual monosaccharide residues¹⁹. The rapid profil-

Figure 3 Normal phase HPLC profile of fluorescently labelled glycans released from human erythrocyte CD59. The elution positions of a ladder of glucose oligomers are shown in the scale above the oligosaccharide profile. This scale is used to assign glucose unit values to the peaks in the chromatogram^{19,20}. The bar chart above the chromatogram has been constructed using the elution positions of standard sugars and can be used to obtain a rapid overview of glycosylation profiles. The hatched bars indicate (i) the elution range for *N*-linked oligomannose sugars (GlcNAc2Man5–9) and (ii) the elution ranges for *N*-linked bi-, tri- and tetra-antennary complex glycans. The open-ended bars indicate that further substitutions (including lactosamine extensions) result in sugars eluting with larger glucose unit values; (iii) the elution ranges for di-, tri- and tetra-saccharides typically associated with *O*-glycans.



ing and simultaneous analysis of the 36 major glycans attached to human erythrocyte CD59 has been carried out in this manner²⁰.

The glycosylation of microgram amounts of a protein can now be determined within 2 to 3 days. For example, the glycans attached to human neutrophil gelatinase B have been simultaneously sequenced using multiple enzyme arrays (Fig. 4a). The purified protein sample may be submitted to hydrazinolysis to release the sugars, or the glycans can be released directly from SDS-PAGE gels, simplifying the purification as illustrated below.

Sugars from protein in SDS gels

A novel 'in-gel release' method has been developed to obtain N-linked oligosaccharides directly from 1-D gels of glycoproteins available only in microgram amounts¹⁷. The pools of charged and neutral sugars can be characterized using the new HPLC technology and neutral and sialylated sugars may also be analysed by MALDI MS. Moreover, the protein remaining in the gel can be used for peptide sequencing following further in-gel digestion with trypsin.

For example, gelatinase B is an inducible metalloproteinase involved in the remodeling of the extracellular matrix. Depending on the producer cell type and the post translational modifications various molecular forms of human gelatinase B have been detected. Purified enzyme preparations of natural gelatinase B are available only in microgram amounts. Purification usually involves substrate affinity chromatography on gelatin-sepharose. However, purified preparations of neutrophil gelatinase B may contain several molecular variants. These cannot be separated by conventional means but can be resolved on SDS gels.

Figure 4b shows a gel of gelatinase B and alongside it the zymogram that indicates which bands contain active enzyme. The 92K band was excised and the glycans released from the gel piece with PNGase F and analysed by MALDI MS and HPLC. The data were consistent with the HPLC analysis of the N-glycans of neutrophil gelatinase B (Fig. 4a).

Developing future applications

This method can be used where limited quantities of material and protein purification problems have previously precluded a detailed glycan analysis. Reducing gels allow the straightforward separation of proteins into their component subunits. For example, IgG can be resolved into heavy and light chains¹⁷ and the surface coat proteins of hepatitis B virus or particle can be separated into three major glycoprotein components (S, M and L)²⁵. Also, when gelatinase B from some cell types is purified on gelatin-Sepharose it is contaminated with its natural inhibitor (TIMP-1)²⁶. This enzyme-inhibitor

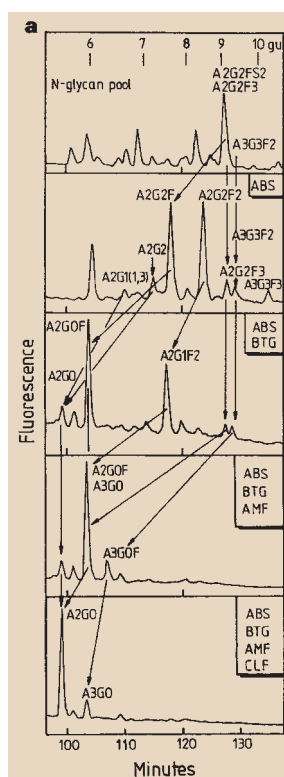
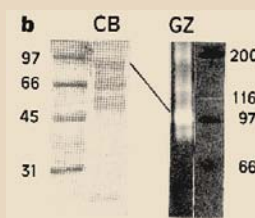


Figure 4a Simultaneous exoglycosidase sequencing of the neutrophil gelatinase B N-glycan pool using the enzyme arrays shown in the panels. Some of the major peaks have been annotated and the arrows indicate the movement of peaks following digestion with the enzyme arrays. The structures of the oligosaccharides are determined from the known incremental values for monosaccharide additions to glycan cores and the specificity of the exoglycosidase enzymes in the arrays. The assignments were consistent with the MALDI MS analysis. (ABS: *Arthrobacter ureafaciens* sialidase, BTG: bovine testis galactosidase, AMF: Almond meal fucosidase, CLF: *Choronia lampas* fucosidase.) b, SDS-PAGE of human neutrophil gelatinase B. Gel electrophoresis was used as the final purification step. Track 1 contains molecular weight markers, track 2 (CB) contains human neutrophil gelatinase B stained with Coomassie blue, track 3 (GZ) shows the result of gelatinase zymography²⁶ used to identify the band containing the full-length active enzyme (92K)²⁷. This band was excised from the gel for glycan profiling using the method described in ref. 17.



complex is stable and SDS-PAGE is used to resolve the two proteins. N-linked oligosaccharides can then be released directly from the gel with PNGase F.

In principle, the method could also be applied to 2-D gels. Glycan analysis will complement protein sequencing in the Proteome Project in which 2-D gels are used for comparing patterns of proteins from cells, tissues or body fluids. Gels of normal and diseased tissue may be compared using computer-controlled imaging techniques, which can identify 'spots' present in one sample but not the other. The possibility of sequencing sugars attached to proteins in such gel 'spots' suggests that this technology may lead to an advance in understanding the role of glycosylation in disease.

State of the art

These developments represent a dramatic advance in the technology available for oligosaccharide analysis. The minimum requirement for a laboratory is for an HPLC system equipped with a normal phase column and a fluorescence detector, a series of glycan standards and a set of endo- and exoglycosidase enzymes. Indeed, many well-founded laboratories may already be in a position to characterize in detail the glycans attached to the glycoproteins they are studying. □

The authors are at the Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK. G.O. is also at Rega Institute for Medical Research, University of Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium.

The analyses referred to in this article were carried out in the Glycobiology Institute using an HPLC system consisting of a Douglas

Scientific DG604 degasser, two Waters 510 pumps, a Waters 717 autoinjector, a Waters temperature control module and a Jasco FP-920 fluorescence detector. The system was controlled via a Waters LAC/E box using Waters Expertease 3.1 software running on a DEC VAX 4000-200 computer. The normal phase, and weak anion exchange HPLC columns (GlycoSep N and GlycoSep C, respectively) were obtained from Oxford GlycoSciences (OGS) UK. Fluorescent labelling with 2-aminobenzamide was carried out using a Signal labelling kit, glycans were released by hydrazine using the GlycoPrep 1000 (OGS) or by PNGase F. Mass spectrometry was carried out using a PerSeptive Biosystems Voyager Elite MALDI TOF mass spectrometer.

- Rudd, P.M. & Dwek, R.A. *Crit. Rev. Biochem. Mol. Biol.* 32, 1–100 (1997).
- Kobata, A. *Anal. Biochem.* 100, 1–14 (1979).
- Edge, C.J. *et al. Proc. Natl. Acad. Sci. USA* 89, 6388–6342 (1992).
- Linden, C. & Lawhead, C.L. *J. Chromatogr.* 105, 125 (1975).
- Palmer, J.K. *Anal. Lett.* 8, 215 (1975).
- Alpert, A.J. *et al. Chromatogr. A* 676, 191–203 (1994).
- Hase, S. *Methods Enzym.* 230, 225–237 (1994).
- Mellis, S.J. & Baenziger, J.U. *Anal. Biochem.* 134, 442–449 (1983).
- Tomiyama, N. *et al. Anal. Biochem.* 193, 90–100 (1991).
- Takahashi, N., Nakagawa, H., Fujikawa, K., Kawamura, Y. & Tomiyama, N. *Anal. Biochem.* 226, 139–146 (1995).
- Kakehi, K. & Honda, S.J. *J. Chromatogr. A* 720, 377–393 (1996).
- Hermentin, P. *et al. Anal. Biochem.* 221, 29–41 (1994).
- Jackson, P. *Anal. Biochem.* 196, 238–244 (1991).
- Lennarz, W.J. & Hart, G.W. eds., *Meth. Enzym., Guide to techniques in glycobiology* 230, 1–520 (Academic, San Diego, California, 1994).
- Dwek, R.A., Edge, C.J., Harvey, D.J., Wormald, M.W. & Parekh, R.B. *Ann. Rev. Biochem.* 62, 65–100 (1993).
- Rudd, P.M. & Dwek, R.A. *Cur. Opin. Biotechnol.* 8, 1–10 (1997).
- Kuster, B., Wheeler, S.F., Hunter, A.P., Dwek, R.A. & Harvey, D.J. *Anal. Biochem.* 250, 82–101 (1997).
- Bigge, J.C., Patel, T.P., Bruce, J.A., Goulding, P.N., Charles, S.M. & Parekh, R.B. *Anal. Biochem.* 230, 229–238 (1995).
- Guile, G.R., Rudd, P.M., Wing, D.R., Prime, S.B. & Dwek, R.A. *Anal. Biochem.* 240, 210–226 (1996).
- Rudd, P.M. *et al. J. Biol. Chem.* 272, 7229–7224 (1997).
- Guile, G.R., Wong, S.Y.C. & Dwek, R.A. *Anal. Biochem.* 222, 231–235 (1994).
- Wormald, M.W. *et al. Biochemistry* 36, 1370–1380 (1997).
- Wing, D.R. *et al. Glycoconjugate J.* 9, 293–301 (1992).
- Iourin, O. *et al. Glycoconjugate J.* 13, 1031–1042 (1996).
- Mehta, A., Lu, X., Block, T., Blumberg, B. & Dwek, R. *Proc. Natl. Acad. Sci. USA* 94, 1822–1827 (1997).
- Opdenakker, G., Masure, S., Proost, P., Billiau, A. & Van Damme, J. *FEBS Letts.* 284, 73–78 (1991).
- Masure, S., Proost, P., Van Damme, J. & Opdenakker, G. *Eur. J. Biochem.* 198, 391–398 (1991).

Acknowledgements: We thank Professor E. M. Southern and Dr Taj Mattu.

This is a mixed bag of new products that includes bibliographic and reference software, chromatography media and columns, a diode array detector, a microplate luminometer and sample heating and cooling units.

XPC knockout mouse

From Taconic

A model for toxicology and drug discovery, exhibiting a deficiency in XPC (xeroderma pigmentosum complementation group C)

These mice are highly susceptible to skin carcinoma after ultraviolet exposure. The mouse XPC gene has extensive sequence homology with the human XPC gene, according to the manufacturer. People with a defective XPC gene exhibit a rare autosomal disease characterized by hypersensitivity of the skin to sunlight and increased risk of skin cancer. The mice are deficient in both alleles of XPC and provide *in vivo* models for studying mechanisms of nucleotide excision repair, for studying ultraviolet exposure on skin carcinogenicity and ultraviolet-induced histopathological skin changes, as well as conducting early stage chemical carcinogenicity screening.

Reader Enquiry No. 100

Luminescence kits

From Labsystems

Three easy-to-use kits expand the applications of this bioluminescent luciferase reaction

The adenosine triphosphate (ATP), luciferase, and cytotoxicity and cell proliferation kits use light emission for measuring concentrations and cell numbers. The cell proliferation and cytotoxicity kit is an alternative to tritiated thymidine, giving results of cell growth and cell death in minutes by measuring ATP present in all metabolically active cells. The luciferase monitoring kit measures transcriptional activity in systems where the firefly luciferase gene is used as a reporter gene. The optimization of reagents allows a detection limit of 0.01 pg of luciferase. Moreover, the assay is said to provide high sensitivity with a stated linearity over five orders of magnitude. Constant-light technology measures ATP on bacterial, plant or animal cells over a concentration range of 10^{-11} to 10^{-6} M. This technology is said to eliminate the need for precise timing.

Reader Enquiry No. 101

Separation and detection

LightPipe Flowcell

From Thermo Separation Products

New optical technology that is used in the UV600LP photodiode array HPLC detector

This flowcell has been designed to increase

the signal-to-noise ratio by extending the length to 50 mm, five times that of standard flowcells. This change is said to heighten the signal by up to 400 per cent without significantly increasing noise. A low-refractive-index polymeric internal surface is used so that no light is absorbed by the flowcell. (Typical short-term noise is better than $\pm 0.3 \times 10^{-5}$ AU cm⁻¹.) To limit baseline noise at longer wavelengths, a 35 W deuterium lamp with a tungsten-halogen source is used, giving sensitivity from 190 to 800 nm. The detector features an optical beam shaper and a specially configured fibre-optic bundle at the aperture, rather than the light-stealing mechanical slit. This is designed to ensure a maximum amount of light in the optimum shape striking the 512-diode array.

Reader Enquiry No. 102

Lightning columns

From Jones Chromatography

A range of columns for LC-MS, LC-MS-MS and fast LC

The columns range from 1 to 10 cm in length with internal diameters of 2.1 mm, 3.0 mm and 4.6 mm. The columns are packed with 3- μ m and 4- μ m inert, base-deactivated Genesis C8 and C18 packing materials. According to the manufacturer, this material exhibits high performance and stability, without the need for mobile phase additives or ion pair reagents. These columns are packed in the Finesse column hardware in lengths of 10, 20, 30, 50 and 100 mm and having internal diameters of 2.1, 3.0, 4.0 and 4.6 mm.

Reader Enquiry No. 103

Instruments

TR717 microplate luminometer

From Tropix

A fast photon-counting microplate system for the detection of bioanalytes

The system operates over a wide range of analyte concentrations (stated to be greater than seven decades), thus reducing the time and effort devoted to diluting samples. According to the manufacturer, it is sensitive down to the low zeptomole level and 'crosstalk' between wells is negligible. The system can read 384-well and 96-well plates, and is compatible with popular robotic

systems. Programmable scan patterns make the reader suitable for high-throughput screening applications, as well. It has a built-in tactile keypad to control all instrument functions and can be used with or without an external PC. An optional heating module provides stable temperature control from 5 °C above ambient to 42 °C. Teflon jet-bellows inject reagents to provide a higher level of accuracy in dispensing



The TR717 fast photon-counting microplate luminometer for the detection of bioanalytes.

reagents and more complete mixing of well contents compared to conventional syringe injectors.

Reader Enquiry No. 104

U-3010 multitasking spectrometer

From Hitachi Scientific Instruments

A new research-grade, double-beam ultraviolet/visible spectrophotometer

Offering improved optical performance and a new Windows95-based user interface, the U-3010 is available in both single monochromator and double monochromator versions. The system uses Windows95 multitasking software to control up to four spectrophotometers simultaneously. New analysis methods can be created or post-run data processing can be carried out while new data are acquired. Data can be transferred directly into an Excel spreadsheet to allow application-specific calculations to be made. A Seya-Namioka-type monochromator is used in this series to improve the resolution performance and reduce stray light. The monochromator contains an astigmatic concave diffraction grating with variable pitch grooves, which corrects for aberrations. The new diffraction grating is purported to reduce stray light to 0.015 per cent transmittance in single monochromator versions and 0.0005 per cent transmittance in double monochromator versions.

Reader Enquiry No. 105

Bibliographic software

Reference Update

From ISI

A system that makes literature searches easier

With articles from more than 1,300 journals, Reference Update help to put the current information at the user's fingertips. A keystroke or the click of a mouse provides complete, up-to-date bibliographic information and author-assigned keywords. Moreover, it allows the user to define the criteria and to create custom search strategies by using browse lists, truncation and synonym tables. The deluxe edition covers some 1,300 publications, whereas the basic edition covers about 400. More than sixty disciplines are represented — the most heavily cited being neurosciences, bio-

chemistry and molecular biology, pharmacology, cell biology and immunology. Access to the weekly issues is available via the Internet, using FTP. References and abstracts can be exported directly to bibliographic management programs.

Reader Enquiry No. 106

Bookends Plus 4.0

From Westing Software

This bibliography/reference management system is now available for electronic delivery

This program can manage references, notes and citations, as well as generate bibliographies in virtually any format. Improved features include Apple Guide support, the ability to design custom in-text citations that can be inserted into a scanned document and configurable punctuation in author-date citations. The program can import rich-text format files, allowing the acquisition of references from any source, leaving intact styled text information, such as italics, bold, Greek font, and the like. Users can format references in HTML and paste them into Web documents. What is more, colleagues with Internet access can actively search Bookends Plus databases from a form on the user's Web page and retrieve the formatted results. The program returns a list of formatted references, each one containing a hypertext link (for example, a link that will retrieve the reference's abstract). Colleagues can also enter references into the user's password-protected databases using the Web.

Reader Enquiry No. 107

EndNote Plus 2.3

From Niles and Associates

An add-in is now available that integrates EndNote into WordPerfect for Windows and Microsoft Word 97

Integration into WordPerfect for Windows 6.1 and 7 and Microsoft Word 97 allows users to have only one document with which to work. Upon selecting 'format bibliography' from WordPerfect's or Word's tools menu, the program will automatically generate a bibliography at the end of the document. The add-in can handle last-minute revisions that are made, such as adding or deleting citations or selecting another bibliographic style. Each time a user selects 'format bibliography', it automatically updates the in-text citations and bibliography in the document.

Reader Enquiry No. 108

Chilling, heating and sealing samples

Plate sealing system

From Porvair Sciences

A microplate sealing system for improved sample integrity and less repetitive strain

With growing concern regarding repetitive strain injury in the workplace, the plate sealer should help alleviate the risk of strain that can occur from having to repetitively hand seal large quantities of microplates, as well as helping to ensure sample integrity. The plate sealer provides an accurate, tight seal on 1-ml and 2-ml deep-well plates and on the Micronic 96-tube rack system. The plate sealer is said to be easy to use, requiring only one operation of the system to seal the plate. Being compact and portable the plate sealer can be easily transported around a laboratory to where it is needed.

Reader Enquiry No. 109

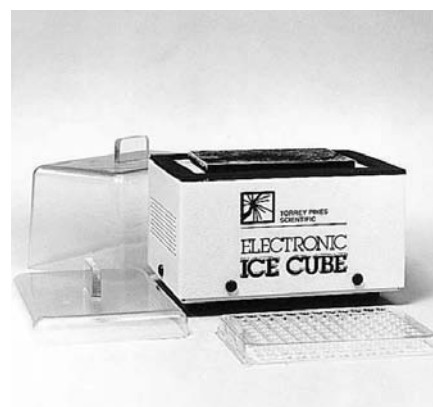
IC10 Electronic Ice Cube

From Torrey Pines Scientific

A compact unit that chills samples to below 0 °C on the bench or in the field

This is a Peltier-driven unit with a chilled plate surface measuring 73 mm × 111 mm. The plate surface is open and accessible for use with robotic systems. It said to be well suited for 96-well assay plates and aluminium blocks designed for 0.2-, 0.5-, and 1.5-µl centrifuge tubes and other samples. Moreover, the unit is small enough to serve as a chilled microscope stage. The chilling module measures 16.5 cm × 12 cm × 8.9 cm (W × D × H) and is supplied with the chiller module, a separate 12-V DC benchtop UL-, CSA- and CE-rated power supply, an AC line cord, and a cable connecting the power supply to the chiller module. An automobile cigarette lighter plug on one end of the cable allows power to be taken from a 12-V DC car cigarette lighter jack for work in the field. The unit has no controls and the unloaded plate surface will go to 0 °C in about two minutes. The maximum differential for the unloaded unit is 30 °C below ambient. Plate covers for 96-well assay plates and larger covers for larger samples are available, as are several aluminium sample blocks.

Reader Enquiry No. 110



Ice Cube chills on the bench or in the field.



Eppendorf's combined sample mixer/incubator.

Thermomixer comfort

From Eppendorf

A unit that mixes and incubates sample tubes

With active countercooling and a temperature range from 13 °C below room temperature to 99 °C, this mixer can be used for a variety of applications. It can mix and incubate samples in four different exchangeable thermoblocks, which can accommodate 24 micro-test tubes of 0.5 ml, 1.5 ml or 2.0 ml, as well as microtitre plates with different well contours. Samples are heated and mixed simultaneously at frequencies between 300 min⁻¹ and 1,400 min⁻¹, and then rapidly cooled down. Each Eppendorf thermoblock also comes with a cooling pack, which ensures rapid cooling down to 0 °C, eliminating the need for an ice bath. Everyday lab routines are made easier by two different programs for selecting heating, mixing and rest phases, as well as subsequent cooling down to room temperature. A short-mix function is also available if extremely short mixing times are required. All operating parameters can be recorded at any time through an RS 232 interface.

Reader Enquiry No. 111

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

**CUSTOM ANTIBODIES
TO PEPTIDES AND PROTEINS**

- Purification • Conjugation
- ELISA Titration
- Hybridoma Development

Also Available

1-800-481-9737

e-mail: antibodies@htibio.com
http://www.htibio.com
fax: (760) 788-9694
P.O. Box 1319
Ramona, CA 92065

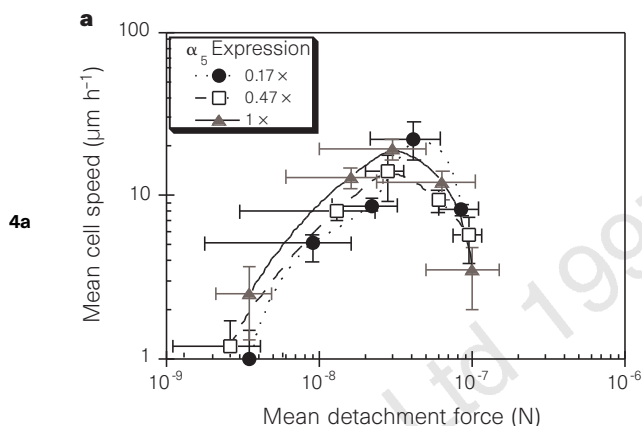
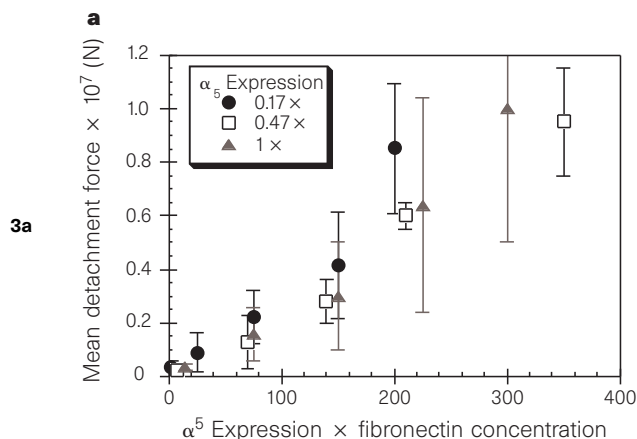
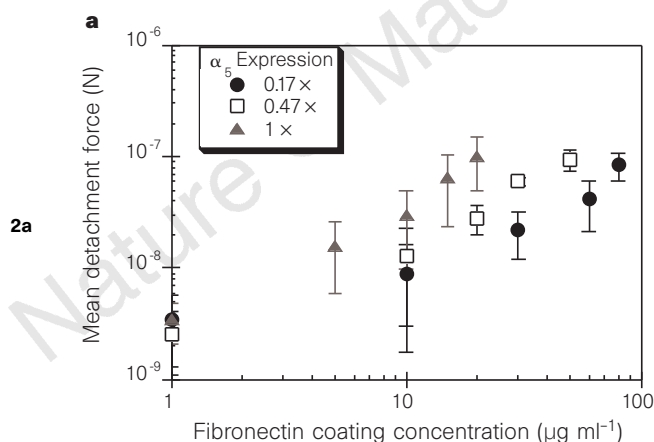
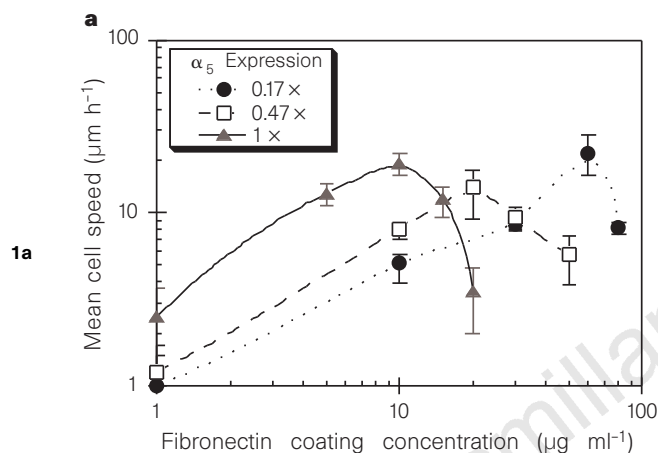
HTI
HTI Bio-Products, Inc.

Integrin–ligand binding properties govern cell migration speed through cell–substratum adhesiveness

Sean P. Palecek, Joseph C. Loftus, Mark H. Ginsberg, Douglas A. Lauffenburger & Alan F. Horwitz

Nature 385, 537–540 (1997)

The triangle symbols in Figs 1a, 2a, 3a and 4a failed to reproduce satisfactorily: the complete figures are reprinted here. □



Total synthesis of the potential anticancer vaccine KH-1 adenocarcinoma antigen

Prashant P. Deshpande & Samuel J. Danishefsky

Nature 387, 164–166 (1997)

In Figs 1 and 2, various hydroxyl groups were incompletely represented. The corrected figures are shown here (Fig. 1, top; Fig. 2, bottom). □

